QUANTITATIVE DATA VALIDATION
(AUTOMATED VISUAL EVALUATIONS)

Anthony John Michael Martin, BSc (Hons.)

A thesis submitted in partial fulfilment of the requirements of De Montfort University for the degree of Doctor of Philosophy

September 1999

De Montfort University

To my parents Fatima and Peter, and my fiancee Celine

A scientist who has formulated a certain hypothesis did not formulate it by chance; it optimally suits his general philosophy in the given domain, his usual way of interpretation, his knowledge and research methodology. He is certainly very anxious to preserve his initial interpretation not only for his own prestige - which is certainly an important factor - but chiefly because it is the hypothesis which is best integrated in the structure of his reasoning. He will be unwilling to give up this first hypothesis because by renouncing it he has to re-evaluate a whole system of conceptions.

Fiscbein and Efraim, 1987.

ABSTRACT

Historically, validation has been performed on a case study basis employing visual evaluations, gradually inspiring confidence through continual application. At present, the method of visual evaluation is the most prevalent form of data analysis, as the brain is the best pattern recognition device known. However, the human visual/perceptual system is a complicated mechanism, prone to many types of physical and psychological influences. Fatigue is a major source of inaccuracy within the results of subjects performing complex visual evaluation tasks. Whilst physical and experiential differences along with age have an enormous bearing on the visual evaluation results of different subjects. It is to this end that automated methods of validation must be developed to produce repeatable, quantitative and objective verification results.

This thesis details the development of the Feature Selective Validation (FSV) method. The FSV method comprises two component measures based on amplitude differences and feature differences. These measures are combined employing a measured level of subjectivity to form an overall assessment of the comparison in question or global difference. The three measures within the FSV method are strengthened by statistical analysis in the form of confidence levels based on amplitude, feature or global discrepancies between compared signals. Highly detailed diagnostic information on the location and magnitude of discrepancies is also made available through the employment of graphical (discrete) representations of the three measures.

The FSV method also benefits from the ability to mirror human perception, whilst producing information which directly relates human variability and the confidence associated with it. The FSV method builds on the common language of engineers and scientists alike, employing categories which relate to human interpretations of comparisons, namely: 'ideal', 'excellent', 'very good', 'good', 'fair', 'poor' and 'extremely poor'.

ACKNOWLEDGEMENTS

My thanks go to my supervisor and mentor Dr. Alistair Duffy for his enduring opinions, encouragement, guidance and continual support.

I gratefully acknowledge the invaluable help of the following people in the preparation of this thesis: Dr. Celine Turner, Paul Cartright, Trevor Benson and Malcolm Woolfson.

Special thanks go to my parents and my fiancee for their support. Without their patience this thesis would not have been possible.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ADM	Amplitude Difference Measure
DNA	Deoxyribonucleic Acid
DSP	Digital Signal Processing
EMC	Electromagnetic Compatibility
FDM	Feature Difference Measure
FSC	Feature Selective Correction
FSV	Feature Selective Validation
GDM	Global Difference Measure
GDT	Global Difference Tolerance
r.f.	Radio Frequency
TLM	Transmission Line Modelling

CHAPTER 1

INTRODUCTION

1. INTRODUCTION

Over recent years, validation/verification has become an integral part of many fields of study. Confidence associated with new technologies and methods of acquiring data sets have evolved through the competent application of visual evaluation studies. Visual evaluation is the foremost method of validation to date, boasting high levels of confidence from the combined assessments of highly skilled subjects (discussed in detail in Section 2.5). However, human variability is a contributing factor to the desire to develop quantitative automated validation methodologies. Fatigue, age and experiential differences between subjects performing visual evaluation tasks all contribute to the phenomenon of human variability[Westcott 1968, Witkin 1954]. These inherent problems associated with the method of visual evaluation do not however undermine the sheer power of the visual/perceptual system within humans to accurately categorise stimuli under ideal conditions. The human brain is the best pattern recognition device known to date[Johnson 1994], and should be acknowledged as such.

Many attempts have been made to move away from the method of visual evaluation, methods such as correlation[Duffy 1994, Woolfson 1995] and reliability factors[van Hove 1997, Zanazzi 1977] being the most successful. These methods have been embarked upon employing the philosophy that automated validation methods should completely replace the process of subjects performing visual evaluations. This philosophy or 'technopoly' is summarised by the American social critic Neil Postman as the widespread view that every ill is a problem which has a potential solution; solutions are provided by technical advances, which are generated by clear, purposeful, disciplined thinking; and the faster the problems are solved the better[Claxton 1997].

To Postman, technopoly is based on

> ***the beliefs that the primary, if not the only goal of human labour and thought is efficiency; that technical calculation is in all respects superior to human judgement; that in fact human judgement cannot be trusted, because it is plagued by laxity, ambiguity, and unnecessary complexity; that subjectivity is an obstacle to clear thinking; that what cannot be measured either does not exist or is of no value; and that the affairs of citizens are best guided and conducted by 'experts'[Postman 1992].***

Automated validation methods such as correlation have been based on this philosophy employing fully automated numerical algorithms without the need for human interaction. However, despite the competent development of such methods, there is to date, no internationally accepted method of fully automated validation which copes with results from all application areas. Current automated validation methods lack discernment in cases where highly detailed diagnostic information is required and lack flexibility due to the removal of human input. Many modern validation methods employ rigid algorithms which do not allow - through the employment of subjective weighting factors - analyses to be tailored to the specific requirements of data sets from diverse application areas. Furthermore, whilst past validation methods have been designed to produce quantitative validation information, it is difficult to interpret with little or no relationship to the common interpretation scale employed by humans performing visual evaluations.

Despite the problems incurred in past designs of automated validation methods, there still remains an increasing requirement for automated validation methods. In order to make progress in the field of automated validation, it is imperative that a new philosophy is followed, not replacing the method of visual evaluation, but automating its mechanisms. The method of visual evaluation must be studied and the underlying mechanics of the scheme must be transferred to powerful modern computers which can replicate the method accurately and efficiently. In transferring this capability to computers the inherent variability incurred between the assessments made by human subjects may be removed, and in-depth diagnostic information based on the quality of complex signal comparisons may be produced. Inevitably, there still remain tasks which can not to date be automated within complicated validation routines, and it is good practice to divide validation schemes between man and machine, allowing humans to engage the tasks which can not easily be replicated by computers. In this way a measured level of flexibility may be associated with otherwise rigid evaluations of compared data sets from widely different application areas.

Automated validation schemes boast the potential advantage of removing the variability from the results of visual validation/verification tasks. Whilst allowing accurate and repeatable results to be obtained more efficiently than the process of human subjects performing identical tasks. The performance of any validation system depends on the variability and diversity of the data to be compared. Whilst the success of any quantitative validation system depends not only on the data and the principles of the system, but the skill and diligence with which these principles are implemented.

1.1 THE IMPORTANCE OF VALIDATION

Within this thesis, validation is the process of checking for critical or subtle defects or imperfections between compared data signals. Deciding what is critical and what is subtle, is a major part of any validation scheme. However, from research in the field of visual evaluations (detailed in Chapter 2), it is clear that two measurements are employed in the evaluation of discrepancies between two stimulus, namely amplitude differences and feature differences. The main purpose of validation is to provide corroborating evidence as to the compliance of technological systems to certain regulations (e.g. EMC, r.f., DSP, optics), and/or the identification and grouping of specific data sets (e.g. finger prints, retina scans, DNA sequences). From these examples alone it is clear that the inherent characteristics of such data will vary immensely. Furthermore, within each of these areas of study, there is a clear trend in the type of validation most commonly employed at present. Highly complex areas of study such as EMC and r.f. rely on the expert eye of dedicated engineers at the expense of both time and cost. Whilst highly repetitive areas of study such as DSP, DNA finger printing and retinal scans rely on systems based on the highly implemented correlation algorithm, in an attempt to produce a measured and repeatable level of feedback from comparison data.

1.1.1 Feedback

Feedback occurs when a system is made aware of the consequences of its actions. Feedback not only gives verification of a systems performance, it allows a system to cope with inconsistent parameters by adjusting its actions in the presence of changing conditions (e.g. noise). Without formal feedback (validation), new technologies and data acquisition methods cannot be fully relied upon to provide solutions that can be used with total confidence. For example, validation of the results of one data acquisition method against another (detailed in Sections 6.1 and 6.2) provides the potential verification and elimination of common assumptions made in both methods which results in the same, but sometimes incorrect, answer. Because of its role in intelligent feedback, quantitative validation can serve as the basis for good research practice in a number of disciplines, including acoustics, linguistics, signal processing, artificial intelligence, electromagnetism and most raw data analysis fields.

1.2 VALIDATION PROTOCOLS

Although it is possible to approach the problem of validation from several view points, such as utilising the knowledge of highly trained scientists in the area under investigation, or applying simple correlation algorithms or more complex reliability functions without the need for human interaction. It is of greater importance to find a balance between necessary human interaction, and computational algorithms which speed up the overall process of validation whilst producing repeatable and accurate results. The fundamental quality which a trained engineer can bring to the validation procedure is the process of subjectively balancing or weighting the core algorithms employed to process data within a specific area of study.

It has been common to see automated validation routines based on single measurements [Duffy 1994, Woolfson 1995, Zanazzi 1977], but it is more appropriate to operate on a multilevel basis[van Hove 1997, Williams 1998]. Within a multilevel validation scheme, individual algorithms are employed to emphasise and extract distinct levels of information embedded in the comparison data sets (detailed in Section 4.2.1). Furthermore, these homogeneous levels must be directly related to the mechanisms inherent in a visual evaluation of results (detailed in Section 2.2). The behaviour of any quantitative validation/verification system depends fundamentally on the extent to which an engineer responds to the information obtained from a comparison. Such dynamic behaviour is difficult to predict[Vernon 1975] and the design of quantitative validation procedures to achieve acceptable response is not a trivial matter. Data validation is not an easy task as there are a number of possible factors which may hinder a comparison of two sets of data. For example, within the field of electromagnetism: experimental noise, the quality of an experimental method, simplifying assumptions made in a numerical model and the Q-factor, position and density of resonant - type features will all complicate a validation procedure.

1.3 VALIDATION CONSTRAINTS

Despite the relatively long period in which visual evaluation has been employed, there is no internationally accepted protocol for validating methods or assessing improvements in new technologies and data acquisition methods. There are a wide variety of potential applications and a single fully automated validation solution to suit all areas would be difficult to conceive and almost impossible to implement. However, it is vital that new methods of data validation are developed and used, in order for new technologies to be employed with total confidence.

One of the most significant problems in the area of validation, is identifying which features are significant, and therefore must be included, those which are helpful, and should be included, and which are of little significance, and should not be included, in a comparison between two signals[van Hove 1994, Williams 1997, Zanazzi 1977].

Research in the area of visual evaluation has indicated that three main measurements[Johnson 1995] are employed during a visual comparison of compared data signals. These three measures, namely: 'atomic', 'relational' and 'positional' difference may be modelled by a series of absolute, first and second order difference equations, which emphasise 'amplitudes', 'trends' and 'features' respectively. However, each of these difference measurements should employ homogeneous regions of the compared data sets. Discrepancies between features must be found within a comparison of high pass filtered data, whilst difference algorithms assessing discrepancies between amplitudes and trends must employ low pass filtered data sets. This methodology allows for the isolation of specific (atomic, relational and positional) discrepancies, allowing true classifications of the types of discrepancies acting upon a comparison of complex data signal sets(detailed in Section 5.4.4).

1.4 PROJECT AIMS

To date, visual evaluation is the most powerful method of data analysis. The brain is the best pattern recognition device known, whilst the human perceptual system allows flexibility within assessments made on the quality of compared signals. This project aims to transfer this capability to an automated validation scheme, improving the speed at which quantitative results may be obtained. In transferring this capability to machines, it is perceived that both the accuracy and reliability of validation results may be increased, allowing a measured level of confidence to be associated with results from a wide cross section of application areas. In this way, new technologies may be validated efficiently allowing rapid prototyping and lower development costs.

Validation is a considerable challenge, a place where the experimental engineer and the numerical engineer must meet. There is no choice; neither alone suffices. This thesis is aimed to help in this meeting.

1.5 OVERVIEW

Chapter 2 details research in the area of visual evaluation, illustrating the mechanisms inherent in the underlying method and the sheer power of the visual/perceptual system within humans. Chapter 3 introduces three current day automated validation schemes, along with results illustrating the ability of these methods to accurately validate complex data sets. Chapter 4 employs the visual evaluation mechanisms researched in Chapter 2 along with the advantages inherent in the automated methods of Chapter 3, in the development of the Feature Selective Validation (FSV) and Feature Selective Correction (FSC) methods. The process requirements of successful validation schemes are discussed, along with a detailed explanation of the development of the FSV and FSC methods. Chapter 5 verifies the theoretical operation of the FSV method, whilst results illustrate the performance of automated validation schemes against a significant amount of information obtained from a survey of visual evaluations among highly skilled subjects. Chapter 6 applies the FSV method to three key application areas, illustrating the immense quantity of information gained during a quantitative evaluation of compared results. Whilst, Chapter 7 details the advantages and disadvantages of visually assessing data sets, along with a discussion on the suitability of the three automated validation methods introduced in Chapter 3. Finally, Chapter 8 discusses the origins and suitability of the mechanisms employed by the FSV method, along with recommendations for further developments within the FSV and FSC methods.

CHAPTER 2

THE HUMAN VISUAL/PERCEPTUAL SYSTEM

2. THE HUMAN VISUAL/PERCEPTUAL SYSTEM

The human visual/perceptual system is the most powerful pattern recognition device known to date. Humans abstract information from visual stimuli in an attempt to produce coherent pictures of their surroundings. However, whilst this method of data extraction and construction is powerful, variabilities between subjects invariably arise. Physical and experiential differences along with the age of subjects all contribute to differences between the 'worlds' we see or perceive. Differences may stem from the way in which critical features are visually extracted from presented stimuli, or the mental maps from which analogies are drawn concerning the nature of the objects under scrutiny. Far from regarding this as a problem, the variability in interpretation by experienced technologists is a real phenomenon underlining the complexity of the compared data sets, and a measured level of confidence may be associated with the combined results of subjects performing identical visual evaluation tasks.

This Chapter describes the main problems associated with employing the results obtained from the process of visually inspecting and assessing graphical data sets, for the purpose of data validation. Variabilities inherent in the process of visual evaluation are detailed along with the mechanisms inherent in the underlying method. Considerable attention is given to the phenomenon of perception and its overriding power to manipulate the brain's 'view' of presented stimuli. Further discussions detail the brain's tendency to categorise stimuli giving each a name, along with the phenomenon of the central category effect. The results presented illustrate the combined evaluations of skilled engineers performing visual evaluation tasks, indicating the level of confidence that may be associated with a comparison of complex data signals. Finally, conclusions are drawn on the suitability of visual evaluation as a robust and accurate method of data validation.

2.1 THEORY

The human brain obtains information through the senses. Within the field of visual evaluation, messages to the brain are provided by the eyes. The brain decodes these messages depending upon the special centres of the brain in question, with most messages transported from the eyes being received in the visual cortex located at the rear of the brain. These messages are sent in the form of electrical stimuli and are interpreted as vision by the brain. The messages described may be viewed as data, which must be interpreted into useful information before an analysis can be made. That is to say, a mental model[Vernon 1975] of the nature of the signals to be analysed must first be constructed from the data provided by vision. Only after this mental model has been built can the process of comparison begin.

2.2 VISUAL SEARCH

The task of visually searching[Eriksen 1990, Krose 1990, Neisser 1970] for patterns and features to construct coherent 'mental pictures' of stimuli can be complicated by the physical, mental and experiential characteristics of subjects[Fozard 1977, Postman 1992, Schneider 1977, Shiffrin 1977]. In its simplest form, visual search may be viewed as the process of extracting critical features from stimuli[Johnson 1995] in an attempt to gain vital information based on the inherent nature of a stimulus' form. An example of this phenomenon is illustrated in Figures 2.1 and 2.2[Hilsenrath 1990]. Figure 2.1 illustrates a sculpture depicting the head of Queen Nefertiti whilst Figure 2.2 depicts the scanpath[Norton 1971a,b, Yarbus 1967] of a subject observing the picture for two minutes.

Figure 2.1: Head of the Egyptian Queen Nefertiti - taken from *"Feature Extraction and Sensitivity Matching in Visual Search"*, in Visual Search, Brogan D, (editor), Taylor and Francis, 1990.

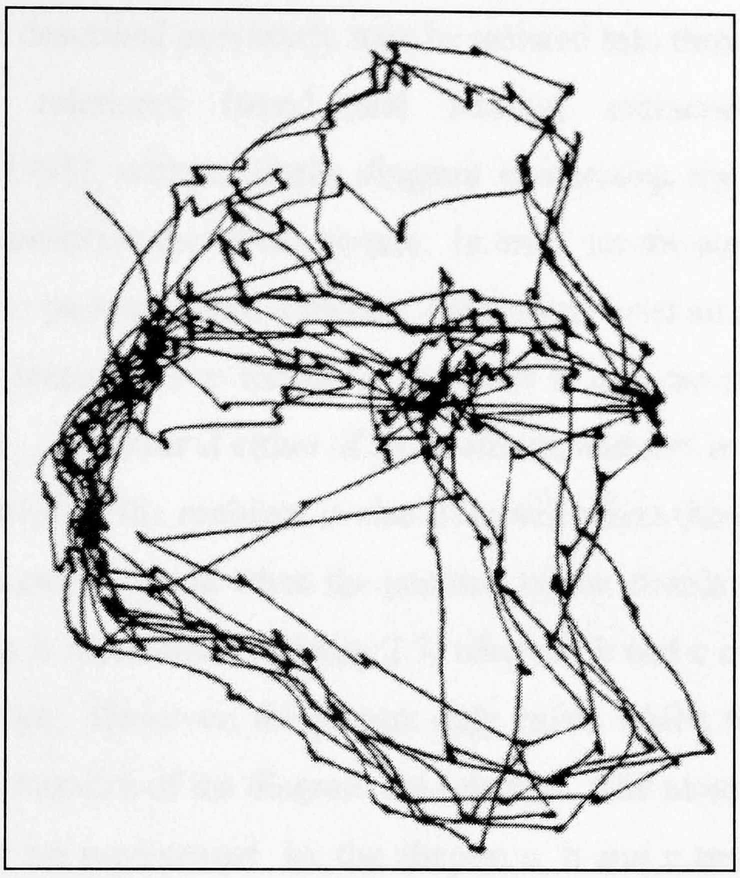

Figure 2.2: Subjects eye movements during free examination of Figure 2.1 - taken from *"Feature Extraction and Sensitivity Matching in Visual Search"*, in Visual Search, Brogan D, (editor), Taylor and Francis, 1990.

The scan trajectory of Figure 2.2, illustrates that only critical areas of the sculpture are studied in an attempt to gain vital information based on the form of the stimulus in the most efficient manner. Significant amounts of attention are recorded in regions of the picture exhibiting intricate features - nose, mouth, ear, eyes and chin. Some attention is given to the overall form of the stimulus where the scan trajectory traces the outline of the full stimulus. Whilst little or no attention is given to areas of the picture exhibiting less relevance to the form of the stimulus - cheek and neck. In scanning areas of high information content, the brain accomplishes a high degree of data reduction at an early stage of visual data acquisition. This reduces the number of features necessary to construct a coherent mental model of the presented stimulus, increasing the efficiency of data extraction and optimising the process of mental stimulus construction.

The scan trajectories described previously may be isolated into three categories, namely: atomic extraction; relational (trend and feature) extraction; and positional extraction[Johnson 1995], with a simple diagram comprising the organisation of the atomic parts which constitute the whole picture. In order for the atomic parts of a figure to represent a specific pattern within a picture, there must exist an inherent relationship between the atomic features. For example, the letter L is constructed employing two atomic features | and _, however if either of these atomic features is removed the letter L is no longer represented. This problem is also observed when the relationship between atomic features is incorrect _| and when the position of the atomic features is changed | _. This phenomenon is illustrated in Figure 2.3, where a, b and c each represent shapes within a simple picture. However, the picture only exists whilst the atomic, relational and positional characteristics of the diagram are retained. The atomic parts employed to construct the picture are represented by the shapes: a, b and c respectively, whilst the relational characteristics of the picture are represented by the three relationships: ab, ac and bc. Furthermore, the positional characteristics of the picture are equated to the co-ordinates of the three atomic features within the figure. If any one of these nine characteristics is changed, this unique picture will no longer exist.

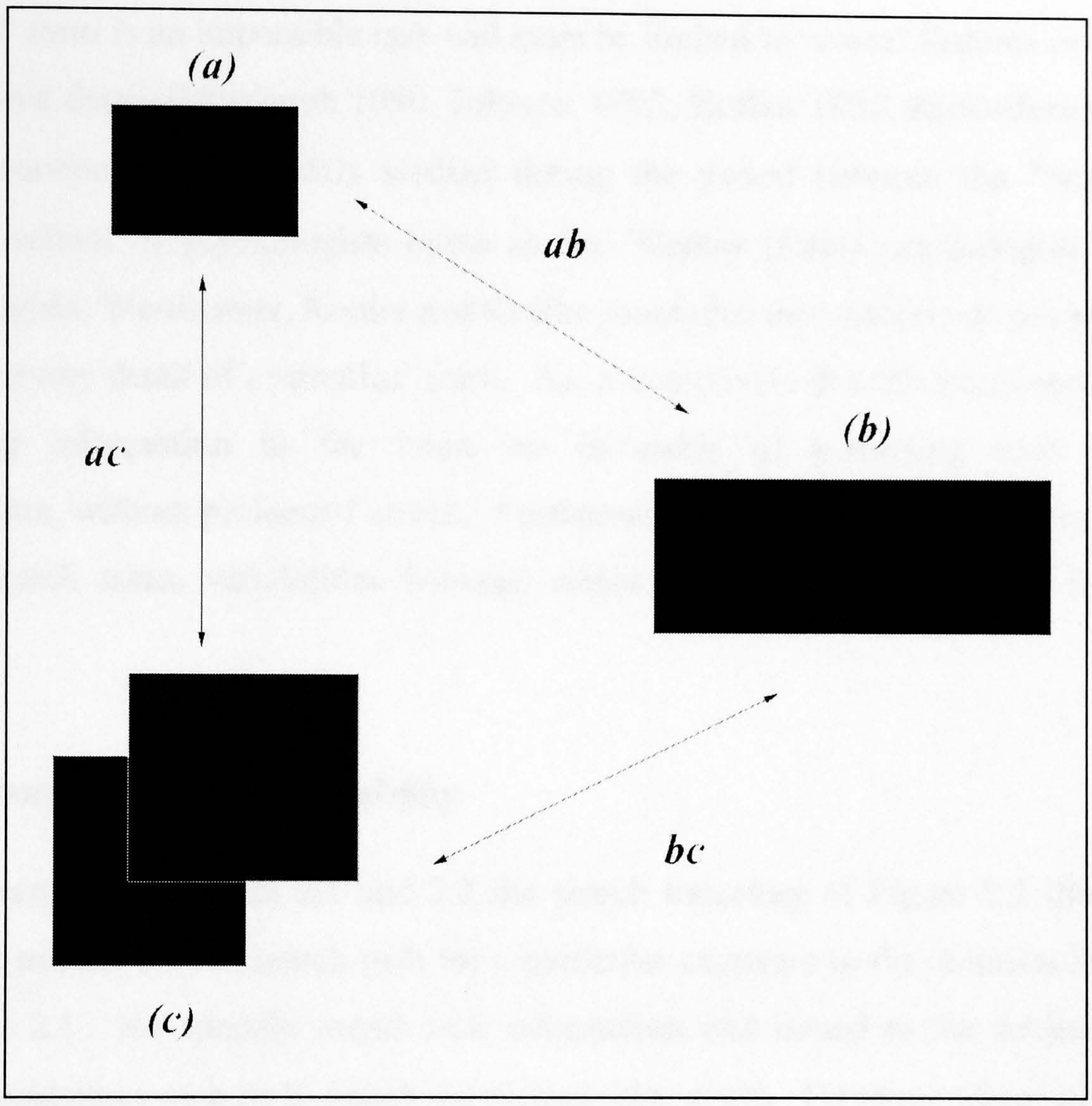

Figure 2.3: Atomic, relational and positional characteristics of a simple diagram

From these analogies, atomic extraction may be viewed as an absolute measure of the intensities within a picture. Relational extraction may be viewed as higher order emphasis routines such as first and second order derivatives. Whilst positional extraction may be viewed as the co-ordinate positions of both the atomic and relational characteristics of a picture. This visual search model (simplified for one dimensional data) forms the basis of the Feature Selective Validation method detailed in Chapter 4.

The information obtained from the visual system assists in the construction of a coherent mental picture or model of the stimulus under investigation[Hilsenrath 1990, Vernon 1971]. As humans, we use visual search to assist in the construction of a coherent model of our surroundings, helping to build an overall picture of the 'world' we live in. However, it is perceived that the extraction of all information denoting a stimulus' form is an impossible task and must be limited to critical features or the most informative details[Hilsenrath 1990, Johnson 1995, Koffka 1935, Kristofferson 1957]. This phenomenon was widely studied during the period between the Ψwars, by a German school of psychologists know as the •***Gestalt*** (form) psychologists. These psychologists: Wertheimer, Kohler and Koffka concluded that subjects do not accurately perceive every detail of a stimulus' form. As, it is probable that the visual mechanisms providing information to the brain are incapable of extracting such complex information, without prolonged search. Furthermore, due to the complexity involved in visual search tasks, variabilities between subjects performing these tasks invariably arise.

2.2.1 Scan/Search Path Variability

In the example of Figures 2.1 and 2.2 the search trajectory of Figure 2.2 illustrates a subject's natural or free search path for a particular exposure to the stimulus illustrated in Figure 2.1. No specific search task information was issued to the subject and no general guidelines on how to search the picture were given. However when subjects are asked specific questions regarding the stimulus they are exposed to, the resulting scanpaths change dramatically. Figure 2.4(a) illustrates 'The Unexpected Visitor of Repin', while Figures 2.4(b) to 2.4(h) illustrate the search patterns of subjects asked specific questions before exposure to the picture[Hilsenrath 1990].

Ψ ***First and second World wars: 1914 - 1918; and 1939 - 1945 respectively.***

• ***Mental stimuli consist of organised wholes (gestalten), not the sum of distinct parts***

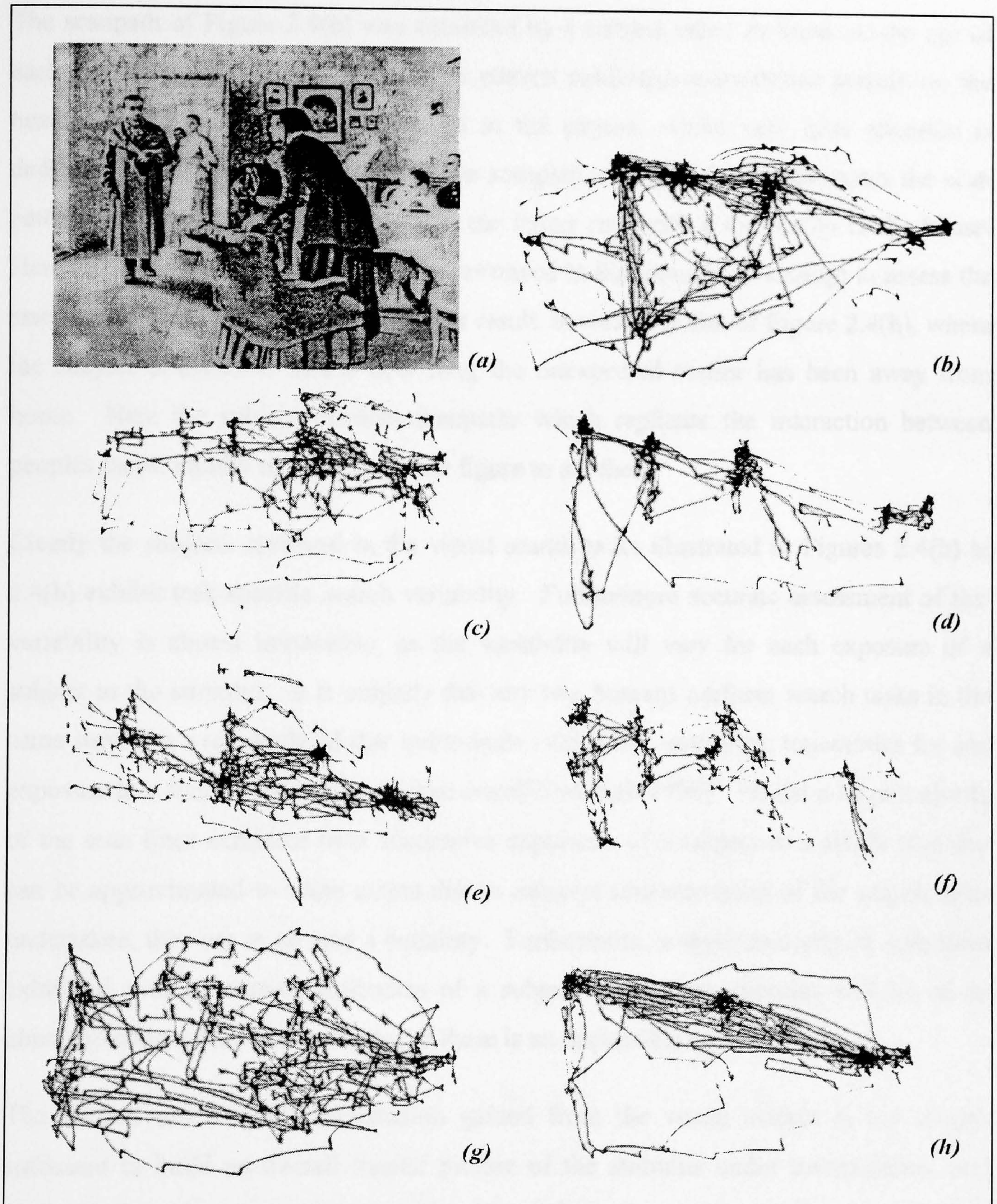

Figure 2.4: Records of seven task specific search paths for same stimulus - taken from *"Feature Extraction and Sensitivity Matching in Visual Search"*, in Visual Search, Brogan D, (editor), Taylor and Francis, 1990.

The scanpath of Figure 2.4(b) was exhibited by a subject asked to focus on the age of each person in the painting. The scan pattern exhibited concentrates mainly on the heads of each of the figures portrayed in the picture, whilst very little attention is dedicated to the rest of the picture. The scanpath of Figure 2.4(g) illustrates the scan pattern of a subject asked to determine the living standard of the family in the house. Here the subject only scans the objects portrayed in the room in an attempt to assess the standard of living. The most spectacular result, however, is that of Figure 2.4(h), where the subject is asked to assess how long the unexpected visitor has been away from home. Here the subject exhibits scanpaths which replicate the interaction between peoples faces, rapidly moving from one figure to another.

Clearly the subjects involved in the visual search tasks illustrated in Figures 2.4(b) to 2.4(h) exhibit task specific search variability. Furthermore accurate assessment of this variability is almost impossible, as the variability will vary for each exposure of a subject to the stimulus. It is unlikely that any two humans perform search tasks in the same way, it is even doubtful that individuals exhibit the same scan trajectories for any exposure to a single stimulus more than once[Hilsenrath 1990]. Whilst a large majority of the scan lines exhibited over successive exposures of a subject to a single stimulus can be approximated to some extent due to inherent characteristics of the search tasks undertaken, they are in no way a certainty. Furthermore, a small minority of scan lines exhibited over successive exposures of a subject to a single stimulus will be of an almost chaotic nature and estimates of these is an impossible task.

The limited and variable information gained from the visual system is not always sufficient to build an overall mental picture of the stimulus under investigation, and further information is required. This extra information is provided in the form of perceptual maps[Covey 1994], allowing rapid construction of coherent mental pictures. Perceptual maps are employed to arrange critical features extracted from stimuli by the visual system in an attempt to construct a detailed and coherent mental model of the stimulus under investigation. Furthermore, these mental maps or paradigms act as templates, filling in the gaps between critical features acquired from the visual system based on a stimulus' form.

2.3 PERCEPTION

The phenomenon of perception allows partial information on 'known' stimuli to be employed in the construction of an overall mental model of a stimulus' form. However, within the stages of stimulus construction, many problems arise due to human or perceptual variabilities between individuals[December 1960, Vernon 1975, Westcott 1968, Witkin 1954]. Due to physical, mental and experiential differences between individuals, the process of both receiving data and constructing an overall mental model of the stimulus may vary enormously. Particular interest is devoted to the phenomenon of personal variability between subjects performing visual evaluation tasks.

2.3.1 Human Variability

Human variability in the field of visual evaluation is a complicated phenomenon. Variations may arise due to variabilities in the scan paths employed by different subjects performing visual search tasks. Furthermore, considerable variabilities stem from the different perceptual maps or paradigms employed by subjects processing information obtained from the visual system[Covey 1994].

A visual comparison may be viewed as a two stage process: a stimulus construction or mental model construction stage; and an analysis stage. Variabilities arise due to individuals interpreting stimuli differently, this may manifest itself at the construction stage of an evaluation and will inevitably exacerbate the problem of assessment variability. That is to say, analysis applied to different mental models of a single stimulus will inevitably provide the seed for different assessments of the comparison signals under investigation.

2.3.1.1 Paradigms

A brain is plastic, it evolves with every new experience it encounters[Claxton 1997, Vernon 1971]. Categories and concepts are instilled from an early age and are updated and added to through both incompetent and competent application of the brain to old and new problems. It is from these categories and concepts, or mental maps, that the process of 'spontaneous analogy' may be called upon[Brain 1941]. These analogies allow past mistakes to be avoided, or new mistakes to be made, whilst developing new and increasingly optimised mental maps until a high level of both competence and confidence is associated with the problem in hand[Carmichael 1932, Herman 1957].

Through the employment of mental maps subjects perceive stimuli on a regular basis and assume that their perception of these stimuli is correct[Vernon 1971]. A subject's perception of stimuli is a balance of evidence, or information gained at that point in time, and the utilisation of information held within the brain as mental maps[Covey 1994]. The brain tunes into certain wavebands of information and evolves these along with its own expanding range of capabilities in order to optimise the understanding of certain stimuli within the 'world' it 'sees'.

To reach a conclusion on the quality of a comparison, an understanding of both the area of study from which the signals were produced and a knowledge of the 'correct' way in which to interpret the data is vital. If this information is not available, or is incorrect, complications will inevitably arise and the outcome may be unreliable. An example of how perceptual maps manipulate the way in which the brain sees stimuli is illustrated in the three sketches of Figures 2.5, 2.6 and 2.7. Figure 2.6 illustrates the visual mask of a young woman, whilst Figure 2.7 depicts the visual mask of an elderly woman. Extensive research[Covey 1994] has shown that the majority of subjects exposed to Figure 2.6 before viewing the stimulus diagram of Figure 2.5, will see or perceive the figure of a young woman. Conversely, subjects exposed to Figure 2.7 before viewing Figure 2.5 will, in the majority, perceive the figure of an elderly woman.

Figure 2.5: Visual stimulus - taken from Covey SR, *"the Seven Habits of Highly Effective People"*, Simon and Schuster, 1994.

Figure 2.6: Visual mask - young woman - taken from Covey SR, *"the Seven Habits of Highly Effective People"*, Simon and Schuster, 1994.

Figure 2.7: Visual mask - elderly woman - taken from Covey SR, *"the Seven Habits of Highly Effective People"*, Simon and Schuster, 1994.

Human perceptual maps may be viewed as being similar to the masks of Figures 2.6 and 2.7, whilst the stimulus under investigation may be viewed as Figure 2.5. From this example different conclusions may be drawn on the nature or characteristics of the stimulus in question, dependant upon the perceptual map or mask employed to construct the overall mental picture. Hence, perceptual maps or paradigms invariably manipulate the visual information received by the eyes and so distort the construction of a subjects mental model of the stimulus they are exposed to[Fernandez 1990, Koffka 1931, Kristofferson 1957]. Furthermore, the example above clearly employs two different paradigms or masks in an attempt to distort a subjects understanding or perception of a stimulus' form. However, within humans it is unlikely that any two subjects possess the same perceptual map of any one stimulus, further exacerbating the problem of human variability.

If it is unlikely that any two humans perceive stimuli in the same way, and it is even doubtful that individuals perceive any one stimulus in the same way more than once, this brings to bear on the problem the question of "who is correct?" This is a difficult question to answer, however it should be noted that a value of confidence may be placed on the combined results of subjects participating in wide scale studies of variability. This process is discussed in detail in Section 2.6 and Chapter 5.

2.3.1.2 Paradigm Shifts

A variety of factors affect the way in which stimuli are perceived, the most obvious of these being the nature of the data being processed, and the skill and motivation of the subject processing the data. The brain itself, due to assumptions based on experience, contributes a great deal to the selectivity in perception. Past experience and training play a large role in an individuals perception of stimuli, particularly in determining the number of differences which can be discriminated among, or told apart. Older subjects perceive stimuli quicker than young children[Haith 1970, Westcott 1968], although older subjects generally possess less visual acuity than their younger counterparts[Vernon 1975]. Many of the variabilities involved in visual evaluations can be reduced by adequate training. Highly skilled engineers exhibit less variability than their lay colleagues, but these variabilities will very rarely disappear entirely even through extensive training[Steinschneider 1990, Unema 1990].

2.4 THE CATEGORY EFFECT

A further component of perception is the brains overriding tendency, whether consciously or unconsciously, to categorise stimuli (the category effect), giving names to each[Cook 1931]. The category effect phenomenon may be viewed as the process of labelling perceived stimuli based on their inherent characteristics or the nature of their form. An extension to this mechanism is the 'centring' effect[Claxton 1997, Koffka 1935, Vernon 1971] within the category effect of perception. For example, many data signals may be different in their characteristics: spatial domain; frequency domain; or time domain, however perception will realise that each is representative of a ***signal***. In this example, each stimulus is labelled differently due to its inherent characteristics (category): ***spatial***; ***frequency***; and ***time***, however a secondary label ***signal*** describing the general nature (central category) of each stimulus is applied. Hence the 'centring' effect allows for the realisation of classical or pure examples of a subjects mental maps.

2.5 THE POWER OF THE VISUAL/PERCEPTUAL SYSTEM

The human visual/perceptual system is an extraordinarily powerful tool, allowing spectacularly efficient abstraction of information from a wide variety of stimuli[Johnson 1995]. Humans employ this system effortlessly every day perceiving that it is a simple process, it is not. Figure 2.8 illustrates a low resolution digitised image of a portrait of Pope Paul III, painted in the sixteenth century by Titan. The section of the picture enclosed by a white rectangle denotes the right eye of the figure. Within this rectangle the boundaries of both the pupil and the white of the eye are very clear to the human eye. However, Figure 2.9 illustrates an enlarged version of this area of the picture, denoting each pixel employed to construct the sub image of Figure 2.8. In Figure 2.9 it is difficult to locate the precise boundaries of either the pupil or the white of the eye, yet when viewing Figure 2.8, these areas are very distinct. It is this uncanny ability to abstract information from visual stimuli that makes the human visual/perceptual system so powerful and difficult to mirror employing machines. However, the visual/perceptual system is the most powerful form of data analysis known at present and in order to make clear progress in the field of validation, this system must be transferred to powerful modern computers.

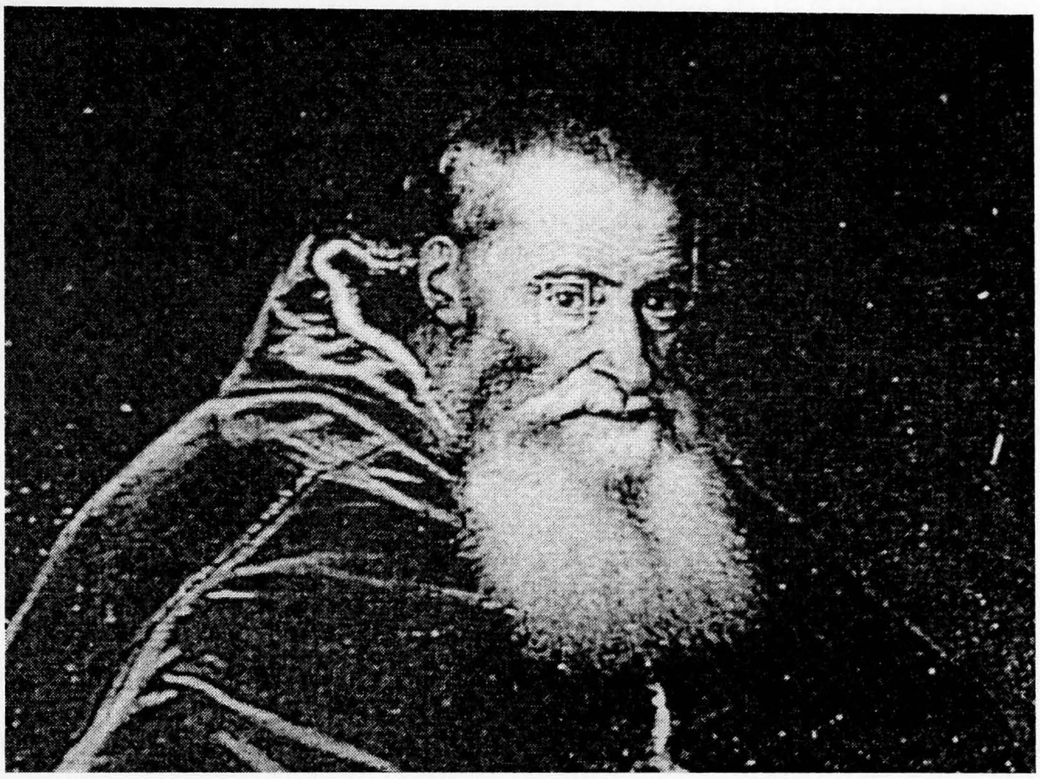

Figure 2.8: Low resolution portrait of Pope Paul III - taken from *"Concepts in Artificial Intelligence"*, Johnson J and Picton P, Butterworth-Heinemann, 1995.

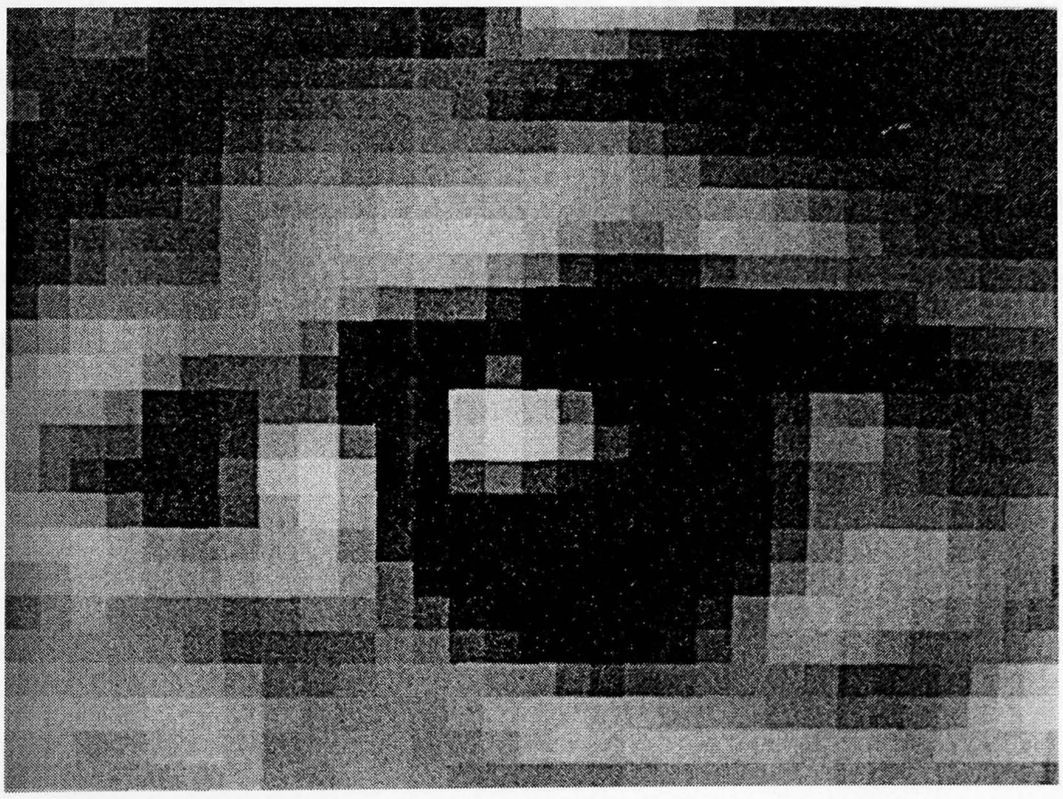

Figure 2.9: Enlarged section of Figure 2.8 - taken from *"Concepts in Artificial Intelligence"*, Johnson J and Picton P, Butterworth-Heinemann, 1995.

2.6 OBTAINING CONFIDENCE FROM COMBINED VISUAL EVALUATION RESULTS

The human visual/perceptual system, whilst powerful, attributes many variabilities to the results of visual evaluations performed by different subjects. In order to gain a level of confidence expressing the quality of a comparison between visual stimuli, it is appropriate to combine the results of a number of subjects performing identical tasks. In this way, variabilities between the results of different subjects may be studied, and a measured level of confidence may be attributed to the nature of compared stimuli.

2.6.1 Method

Figure 2.10 illustrates a comparison of complex data signals, namely $I_{SET1}(f)$ and $I_{SET2}(f)$. Nineteen subjects participated in the experiment in an attempt to visually analyse the comparison of Figure 2.10. The task required each subject to visually assess the comparison, placing it in one of seven quality bands or categories, namely: 'ideal', 'excellent', 'very good', 'good', fair', 'poor' or 'extremely poor'. Information on the general procedure employed in acquiring the comparison sets was not specified. Examples of the experiment were not included in the general task information and no explanation of the meaning of each category was specified, mirroring the environment employed during Hilsenrath's Nefertiti experiment detailed in Section 2.2. Furthermore, the twenty subjects participating in the experiment were trained engineers and scientists. In employing this filter, unnecessary and inappropriate variabilities between assessment results were minimised.

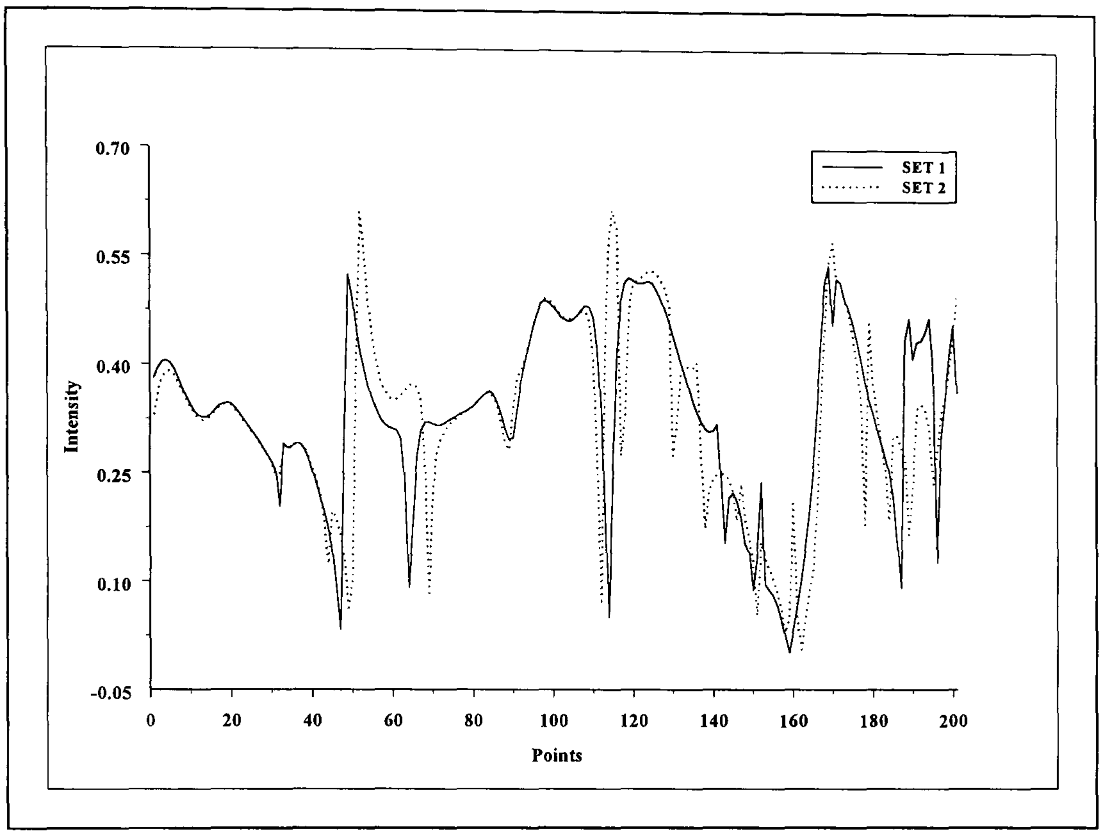

Figure 2.10: Data Sets - $I_{SET1}(f)/I_{SET2}(f)$

2.6.2 Results

Results from this study were processed, with each quality band given a value expressing the percentage of subjects selecting that category to indicate the quality of the comparison illustrated in Figure 2.10. Figure 2.11 illustrates the results from this study.

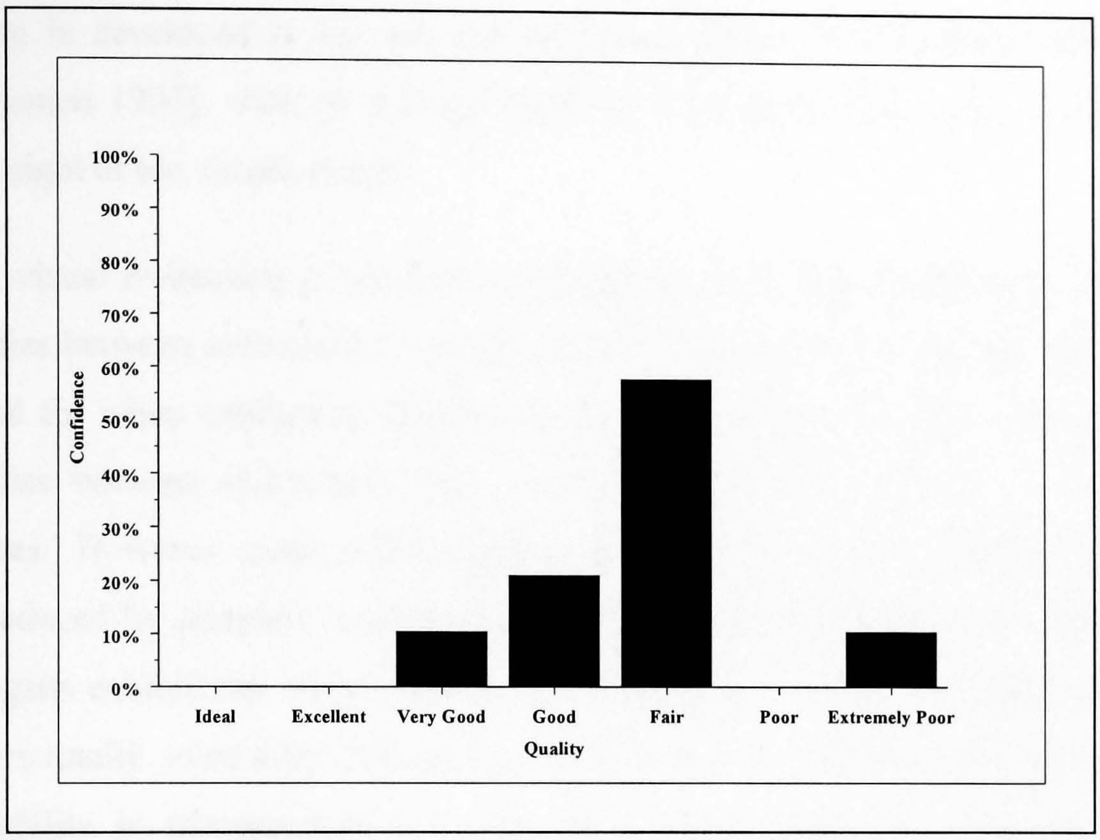

Figure 2.11: Combined visual evaluation results

The results of Figure 2.11 indicate that whilst all subjects associated the quality of the comparison with one of the seven possible categories, the assessments were variable, indicated by widely dispersed confidence levels. However, from the combined results illustrated in Figure 2.11, a measured level of confidence may be associated with the major category chosen to describe the quality of the comparison under investigation, and a valid argument may be presented that the comparison is of 'fair' quality.

2.7 CHAPTER SUMMARY

The method of visual evaluation may be viewed as a two stage process. The first stage involves the recognition of patterns or the extraction of information from a stimulus' form. This process employs the visual search system and perceptual maps individually honed to optimise the accurate construction of mental models. The second stage involves the analysis of several mental stimulus models in an attempt to distinguish differences or similarities between their perceived forms. This process employs heuristics (rules) based on the application area in question and perceptual maps honed to an individuals concept of differentiating between stimuli. The phenomenon of visual evaluation is developed at an early age in humans[Ames 1953, Ghent 1956, Granit 1921, Terman 1937], with an average child of four years of age able to distinguish between eight of ten simple shapes.

To date, visual evaluation is the most prevalent form of data comparison. However, variabilities between individuals' own perceptions is common and a factor that must be accounted for when employing information from a variety of sources. The causes of variabilities between individuals may arise due to physical, mental or experiential differences. However, many of the problems involved in the field of visual evaluation can be reduced by adequate training, mitigating experiential differences. Experienced technologists exhibit less variability than their lay colleagues but this variability rarely disappears totally, even after extensive training. Far from regarding this as a problem, the variability in interpretation by experienced technologists is a real phenomenon underlining the complexity of the comparison data and should be something which an automated validation scheme can reflect.

The results presented illustrate large variations between the category effects of highly skilled engineers performing visual evaluations. These results illustrate that while human variability is a common factor within the field of visual evaluations, levels of confidence can be associated with the combined results of subjects performing these tasks. However, performing visual evaluation tasks on large sets of potentially complex data is a time consuming process requiring high levels of attention[Berylne 1960, Lindsley 1957, Venables 1967] over long periods of time. It is clearly essential that the powerful mechanisms employed by subjects performing visual evaluations are transferred to modern computers. In this way, the problems of both fatigue and assessment variability may be removed from validation results.

From the study of visual search mechanisms detailed in Section 2.2, automated validation routines must possess the ability to mirror the visual emphasis placed upon critical areas of the stimuli under investigation along with their coordinate positions within a given structure. In order to produce this information, both the absolute (amplitude) and relational (feature) properties of a given stimuli must be evaluated. Furthermore, to allow flexibility within otherwise rigid evaluations of discrepancies between two stimuli a measured level of subjectivity must be allowed in the weighting of either amplitudes or features within the overall validation scheme. Whilst chosen by the subject performing the validation task, this level of objective subjectivity is a measurable quantity which may be recorded alongside the validation results. In mirroring these mechanisms inherent in a visual evaluation between two stimuli, automated validation schemes of the future will allow quantitative assessments of data sets from a wide cross section of application areas, whilst reliably producing repeatable and recordable validation results (this is persued further in Chapter 4).

CHAPTER 3

AUTOMATED VALIDATION - CURRENT TECHNIQUES

3. AUTOMATED VALIDATION - CURRENT TECHNIQUES

The comparison of graphical data is a widespread aspect of many disciplines within science, engineering and technology; whether for validating complex data signals or hypotheses, or assessing design changes. Experienced technologists perform visual evaluations in complex application areas at the expense of rapid processing and cost effectiveness. In other areas of study, such as signal processing, where the processes involved are simple and highly repetitive, correlation or reliability factors may be employed. Correlation and reliability factors are techniques in widespread use to quantify the level of agreement or dissimilarity between sets of results.

This chapter presents the methodologies behind current validation techniques in widespread use within the engineering and scientific fraternity. Section 3.1 describes the basic operation of classical correlation algorithms, along with an introduction to some of the key problems related to similarity or multiplication methods. Section 3.2 introduces two methods of validation in wide scale use in the field of X-ray crystallography namely: Zanazzi and Jona; and van Hove reliability factors. Furthermore, a modification to the Van Hove technique which extends the methods ability to produce discrete results is proposed. These methods of data signal validation illustrate the possible advantages in employing pre-emphasis filters and difference equations in the quest for reliable, repeatable and informative validation schemes. Finally Section 3.3 describes the main advantages and disadvantages associated with each of the current automated validation methods. Whilst conclusions are drawn on the suitability of current automated validation methods and the direction in which further advancements in automated validation techniques will be made.

3.1 CORRELATION

One step towards a systematic, objective and robust validation method is the implementation of correlelograms[Duffy 1994, Woolfson 1995]. Correlelograms provide a view of the overall level of agreement between compared signals, employing a measure of similarity. Correlation requires little computational power, whilst providing 'best fit' global figures of merit for successive shifts between data signals. Historically, correlation has been employed in diverse application areas where high speed validation is required, however, modern day correlation techniques are seeing wider application in such areas as: EMC, r.f. and DNA fingerprint analysis. Within areas such as EMC and r.f., correlation is employed as a feedback factor in the optimisation of experimental/modelling procedures and the validation of hypothesis.

3.1.1 Classical Correlation Measures

An evaluation of variance, mean and mean square provide no information about the frequency content of a signal; also, they do not uniquely categorise a particular signal, as in general, a number of signals may share the same mean or mean square values. Correlation overcomes the first of these limitations, but falls short of the second and most major problem of uniquely identifying a signal. Correlation between two signals $I_{SET1}(f)$ and $I_{SET2}(f)$ is normally in the general form of:

$$R(\tau) = \sum_{f_{\min}}^{f_{\max}} I_{SET1}(f) I_{SET2}(f + \tau) \qquad \textbf{(3.1)}$$

where

$$-(f_{\max} - f_{\min}) < \tau < (f_{\max} - f_{\min})$$

where $I_{SET2}(f+\tau)$ denotes a time shifted version of function $I_{SET2}(f)$ and τ denotes shift.

Hence Equation 3.1 and more specifically correlation is the multiplication of a function $I_{SET1}(f)$ with a shifted version of a second function $I_{SET2}(f)$. The result is integrated over the full spectrum of the compared signals f_{min} to f_{max}. Which yields an instantaneous value R_τ for the correlation response $R(\tau)$ corresponding to the shift employed τ. Where $R(\tau)$ denotes a set of values for all possible shifts.

3.1.1.1 Auto-correlation

Auto-correlation is the correlation of a signal $I_{SET1}(f)$ with itself $I_{SET1}(f)$. This provides a measure to which the future value of a signal can be deduced, which is very closely related to the energy spectrum of the signal itself. This is denoted by $R_{11}(\tau)$, given by:

$$R_{11}(\tau) = \sum_{f_{\min}}^{f_{\max}} I_{SET1}(f) I_{SET1}(f+\tau) \tag{3.2}$$

where

$$-(f_{\max} - f_{\min}) < \tau < (f_{\max} - f_{\min})$$

This is often denoted by ✡, such that $R_{11}(\tau) = I_{SET1}(f)$ ✡ $I_{SET1}(f)$.

The shift ($\tau = 1$) illustrated in Figure 3.1 produces a single point on the response curve $R_{11}(\tau)$. Repetition of this procedure for all possible values of τ allows for the complete function $R_{11}(\tau)$ to be obtained. This is simply the shifting of $I_{SET1}(f)$ over itself in both the left and right direction, whilst plotting the area obtained at each shift as $R_{11}(\tau)$. Furthermore, it should be noted that the function $R_{11}(\tau)$ is always symmetrical and $R_{11}(0)$ represents the total energy contained in the signal.

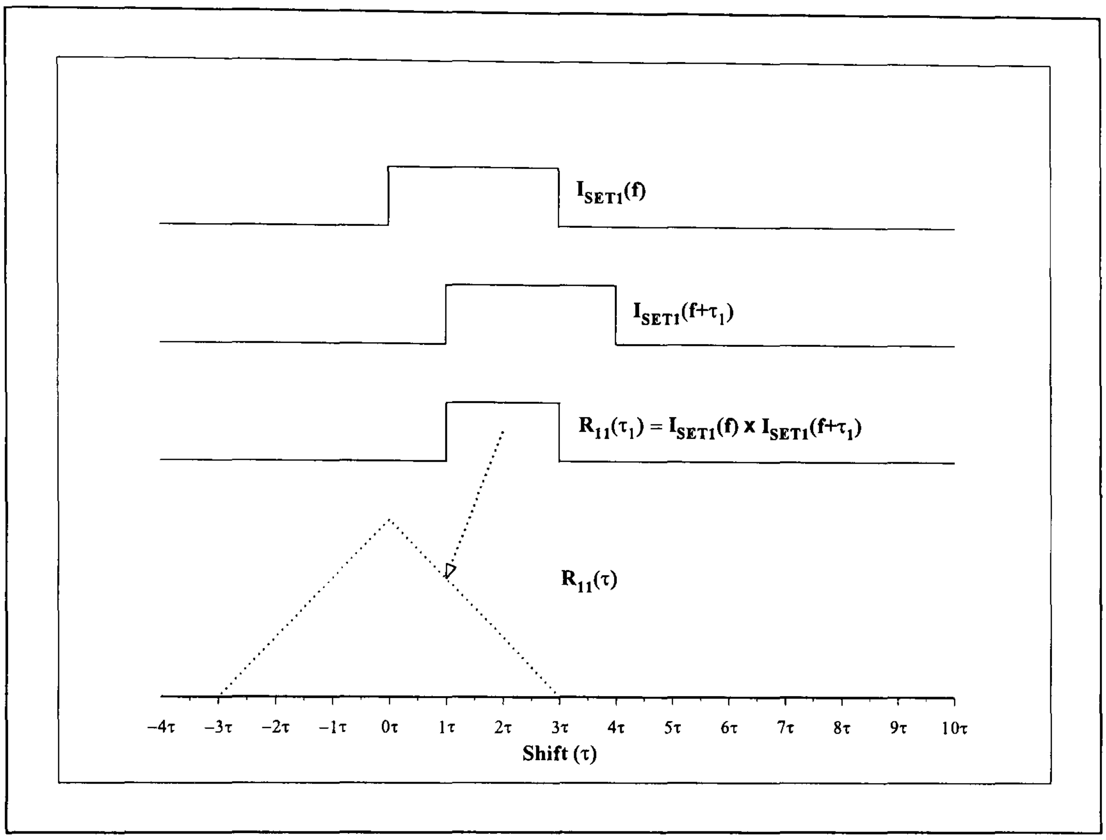

Figure 3.1: Auto correlation $R_{11}(\tau)$ - $I_{SET1}(f)$

3.1.1.2 Cross-correlation

Cross-correlation is a measure of similarity between two signals $I_{SET1}(f)$ and $I_{SET2}(f)$, given by:

$$R_{12}(\tau) = \sum_{f_{min}}^{f_{max}} I_{SET1}(f) I_{SET2}(f+\tau) \tag{3.3}$$

where

$$-(f_{max} - f_{min}) < \tau < (f_{max} - f_{min})$$

or more compactly $I_{SET1}(f) \star I_{SET2}(f)$.

Figure 3.2 illustrates the principle of cross-correlation in detail. One function $I_{SET2}(f)$ is shifted to the right and left and the resulting areas are evaluated, giving the correlation response $R_{12}(\tau)$.

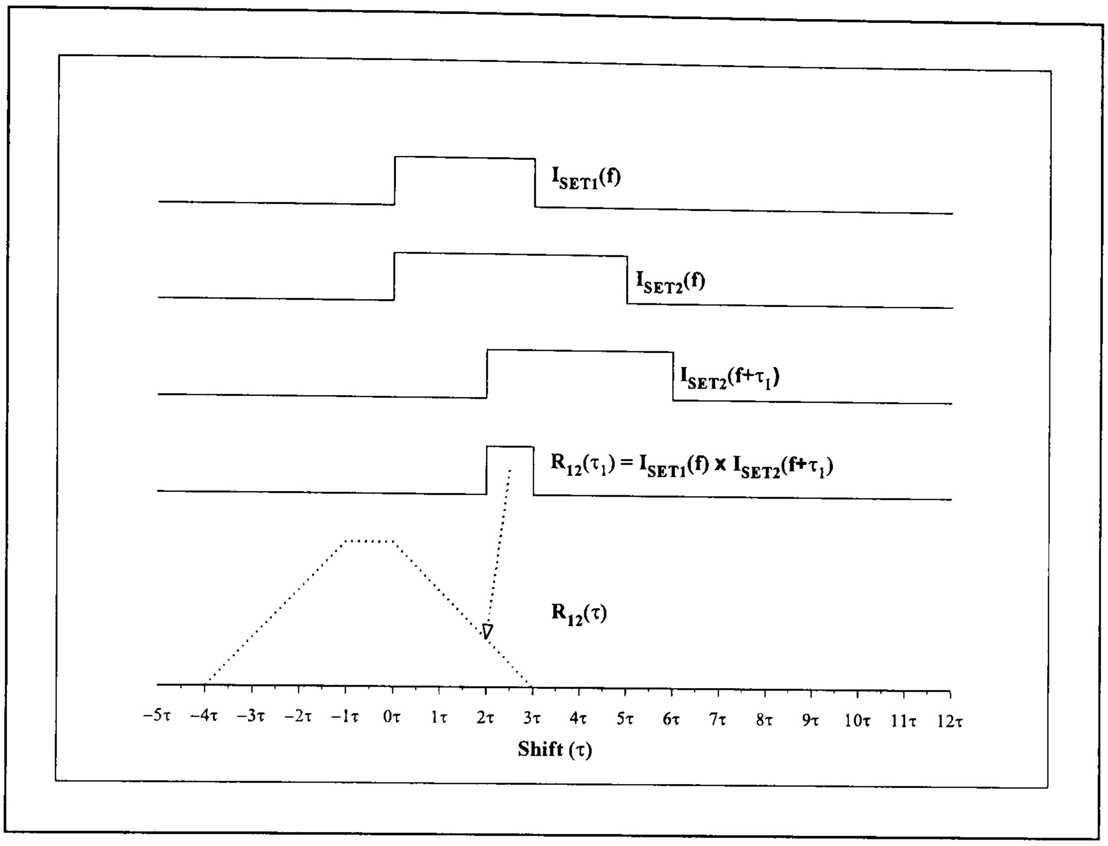

Figure 3.2: Cross correlation $R_{12}(\tau)$ - $I_{SET1}(f)/I_{SET2}(f)$

It should be noted that correlelograms regard all parts of the compared signals as being equally important; there is no latent weighting for feature amplitudes, their positions and the behaviour of the system between these features (where feature denotes a peak, trough or inflexion). This in turn produces a global figure of merit indicating the quality of a comparison.

3.1.2 Extended Correlation

A more specific set of algorithms[Duffy 1994, Woolfson 1995] for computing the correlation of two signals $I_{SET1}(f)$ and $I_{SET2}(f)$ illustrated in Figure 3.3 are given in Equations 3.4 - 3.7. The first of these algorithms denotes the normalised auto-correlation (Equation 3.4) of two signals, and is employed in the assessment of both the RMS symmetry and RMS difference between compared signal sets.

$$R_{11}(\tau) = \frac{\sum_{f_{\min}}^{f_{\max}} I_{SET1}(f) I_{SET1}(f+\tau)}{\sqrt{\sum_{f_{\min}}^{f_{\max}} (I_{SET1}(f))^2 \sum_{f_{\min}}^{f_{\max}} (I_{SET1}(f))^2}} \qquad \textbf{(3.4)}$$

Furthermore the normalised cross-correlation of two signals $I_{SET1}(f)$ and $I_{SET2}(f)$ represented by N samples is given by:

$$R_{12}(\tau) = \frac{\sum_{f_{\min}}^{f_{\max}} I_{SET1}(f) I_{SET2}(f+\tau)}{\sqrt{\sum_{f_{\min}}^{f_{\max}} (I_{SET1}(f))^2 \sum_{f_{\min}}^{f_{\max}} (I_{SET2}(f))^2}} \qquad \textbf{(3.5)}$$

where

$$-\left(f_{\max} - f_{\min}\right) \le \tau \le \left(f_{\max} - f_{\min}\right)$$

$$\tau_{(n)} = n\frac{f_{\max} - f_{\min}}{N}$$

where n denotes the instantaneous sample position within the full compliment of samples N representing the signals.

Figure 3.3 illustrates a comparison of complex data signals $I_{SET1}(f)$ and $I_{SET2}(f)$, whilst Figure 3.4 illustrates the auto-correlation response $R_{11}(\tau)$ obtained employing data signal $I_{SET1}(f)$ and Equation 3.4. Furthermore, Figure 3.5 illustrates the cross-correlation response $R_{12}(\tau)$ obtained employing the comparison of Figure 3.3 and Equation 3.5. The maximum of the cross-correlation response $R_{12}(\tau)$ is referred to as the cross-correlation coefficient and in the comparison illustrated in Figure 3.3 is equated to 0.965 located at zero shift or $R_{12}(0)$. This indicates that the comparison of $I_{SET1}(f)$ and $I_{SET2}(f)$ degrades as successive shifts are applied to $I_{SET2}(f)$. It should be noted that the inclusion of the normalisation factors in Equations 3.4 and 3.5 produces a maximum possible correlation coefficient of 1 for $R_{11}(\tau)$ and $R_{12}(\tau)$ if the two signals $I_{SET1}(f)$ and $I_{SET2}(f)$ are identical.

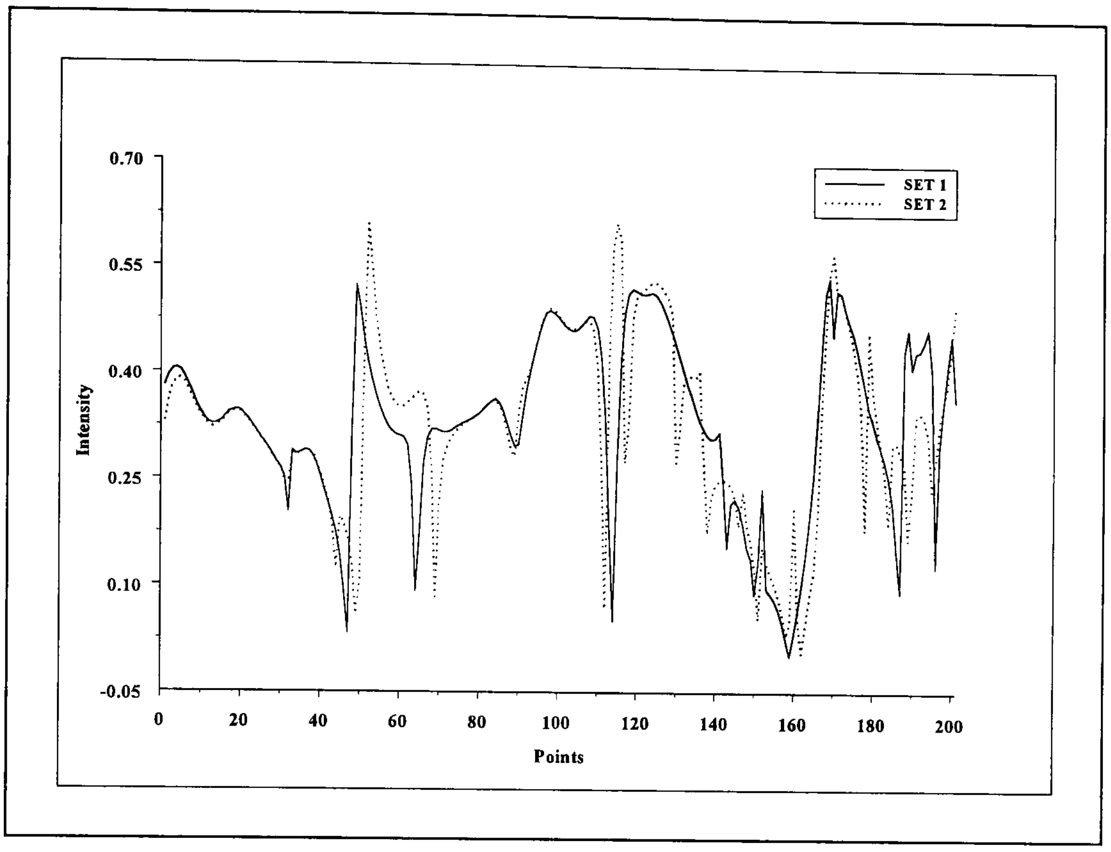

Figure 3.3: Data Sets - $I_{SET1}(f)/I_{SET2}(f)$

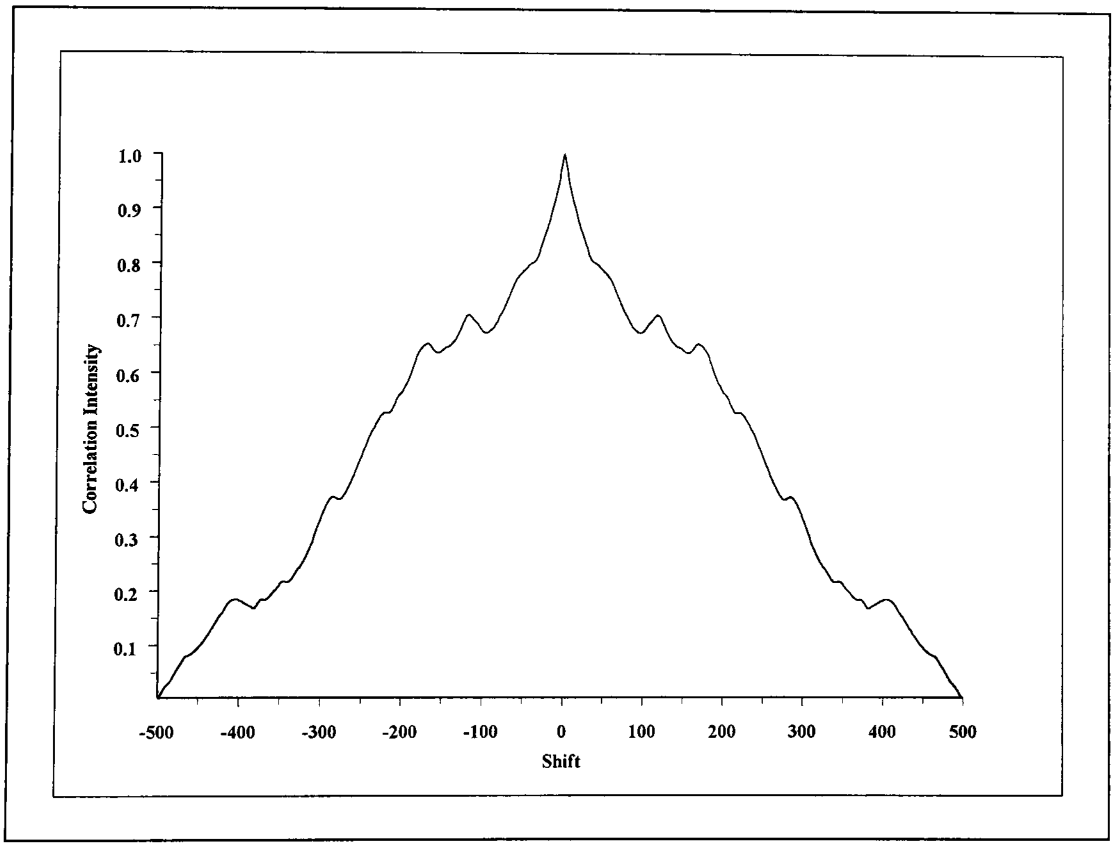

Figure 3.4: Auto-correlation response - $I_{SET1}(f)/I_{SET2}(f)$

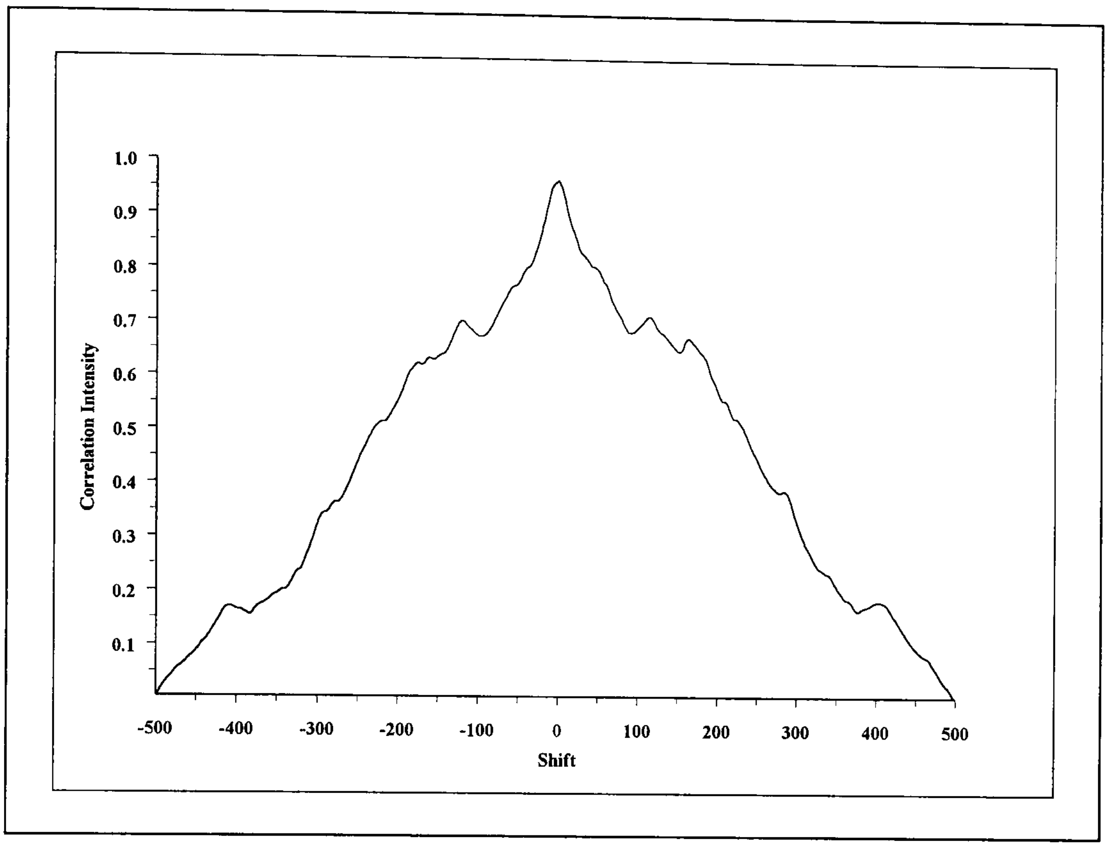

Figure 3.5: Cross-correlation response - $I_{SET1}(f)/I_{SET2}(f)$

In general, when $I_{SET1}(f) \neq I_{SET2}(f)$, the asymmetry of the cross-correlation function, $R_{12}(\tau)$, and its difference from the auto-correlation function $R_{11}(\tau)$, of $I_{SET1}(f)$, provide information about the closeness of $I_{SET1}(f)$ and $I_{SET2}(f)$.

Historically it has been proposed that values derived from correlation functions may support visual comparisons performed by practising engineers. For this purpose three statistical functions were implemented[Duffy 1994, Woolfson 1995] - based on the auto-correlation and cross-correlation functions described previously - to evaluate the similarities between the two response curves, $I_{SET1}(f)$ and $I_{SET2}(f)$. This is achieved employing three measures: the maximum value of the cross-correlation function; the RMS symmetry of the resulting cross-correlation function; and the RMS difference between the cross-correlation and the auto-correlation functions of a comparison. The three extended correlation measures are given by:

1. ***Unity minus the maximum component of the cross-correlation function, (1 - $R_{12}(\tau)_{max}$). As the value of this measure decreases towards zero, the similarity between compared signals increases.***

2. S_{rms}. The degree of asymmetry of $R_{12}(\tau)$.

$$S_{rms} = \sqrt{\frac{2}{\tau_{\max}} \sum_{\tau_{\min}}^{(\tau_{\max}/2)-1} (R_{12}(\tau) - R_{12}(\tau_{\max} - \tau))^2} \tag{3.6}$$

where

$$-\left(f_{\max} - f_{\min}\right) \leq \tau \leq \left(f_{\max} - f_{\min}\right)$$

$$\tau_{(n)} = n\frac{f_{\max} - f_{\min}}{N}$$

$$\tau_{\max} = f_{\max} - f_{\min}$$

where n denotes the instantaneous sample position within the full compliment of samples N representing the signals and τ_{max} and τ_{min} denote the maximum and minimum shifts applied to $I_{SET2}(f)$ respectively.

If $I_{SET1}(f) = I_{SET2}(f)$, then, from the symmetry of cross-correlation function, $S_{rms} = 0$. As S_{rms} decreases towards zero, the similarity between compared signals increases.

3. D_{rms}. The RMS difference between the auto correlation function of the reference data and the cross correlation function of the reference and comparison data sets.

$$D_{rms} = \sqrt{\frac{1}{\tau_{\max}} \sum_{\tau_{\min}}^{\tau_{\max}} (R_{11}(\tau) - R_{12}(\tau))^2} \tag{3.7}$$

where

$$-\left(f_{\max} - f_{\min}\right) \leq \tau \leq \left(f_{\max} - f_{\min}\right)$$

$$\tau_{(n)} = n\frac{f_{\max} - f_{\min}}{N}$$

$$\tau_{\max} = f_{\max} - f_{\min}$$

where n denotes the instantaneous sample position within the full compliment of samples N representing the signals and τ_{max} and τ_{min} denote the maximum and minimum shifts applied to $I_{SET2}(f)$ respectively.

As D_{rms} decreases towards zero, the similarity between compared signals increases.

Table 3.1 indicates results obtained employing the correlation measures described previously applied to the comparison between $I_{SET1}(f)$ and $I_{SET2}(f)$ illustrated in Figure 3.3.

R_{12}	S_{rms}	D_{rms}
0.035	***0.011***	***0.146***

Table 3.1: Correlation results - $I_{SET1}(f)/I_{SET2}(f)$

The use of these three related measurements avoids the potential problem of relying on a single measure to determine the similarity between two data sets in cases where several criteria must be assessed to find the overall similarity. This is accomplished by providing information on different aspects of the data sets under investigation in a controlled manner. However, although the correlation scheme is separated into individual measures identifying the similarities between compared signals, these isolated measures are mathematical abstractions and are not based on the mechanisms employed during a visual comparison of results. Specific measures realising atomic, relational or positional similarities/differences are not employed, and no qualitative or meaningful scales may be applied to the values obtained. Furthermore, correlation does not lend itself to the extraction of a discrete, or point-by-point analysis of compared signals.

3.1.3 Discrete Analysis

In order for automated validation methodologies to be embraced by those who use them, a level of understanding based on the mechanics of the methods must first be demonstrated. For any validation method to be successful it must produce highly detailed diagnostic information regarding the magnitude and location of discrepancies between compared data signals. To overcome these difficulties, the validation method in question must possess the ability to produce a discrete or point-by-point analysis of a comparison of data signals. This step offers to engineers and scientists alike, a detailed map on both the locations and magnitudes of major errors acting upon a comparison, allowing precise feedback for the data acquisition method under investigation. Similarity measures such as correlation do not however lend themselves to an accurate analysis of discrete samples, due to the inherent characteristics associated with the normalisation method employed in a comparison between data signals. This is illustrated in Examples 3.1 and 3.2 where three data sets $I_1(f)$, $I_2(f)$ and $I_3(f)$ are correlated with no shifts applied.

Example 3.1: Discrete correlation analysis - $I_1(f)/I_2(f)$.

Figure 3.6 illustrates the comparison of two data signals $I_1(f)$ and $I_2(f)$. Major discrepancies are observed at five discrete points, namely: 2, 6, 7, 8 and 11. A discrete analysis, should at a minimum, indicate that discrepancies exist at these locations. Furthermore, validation information should indicate that the largest discrepancy is located at position 11.

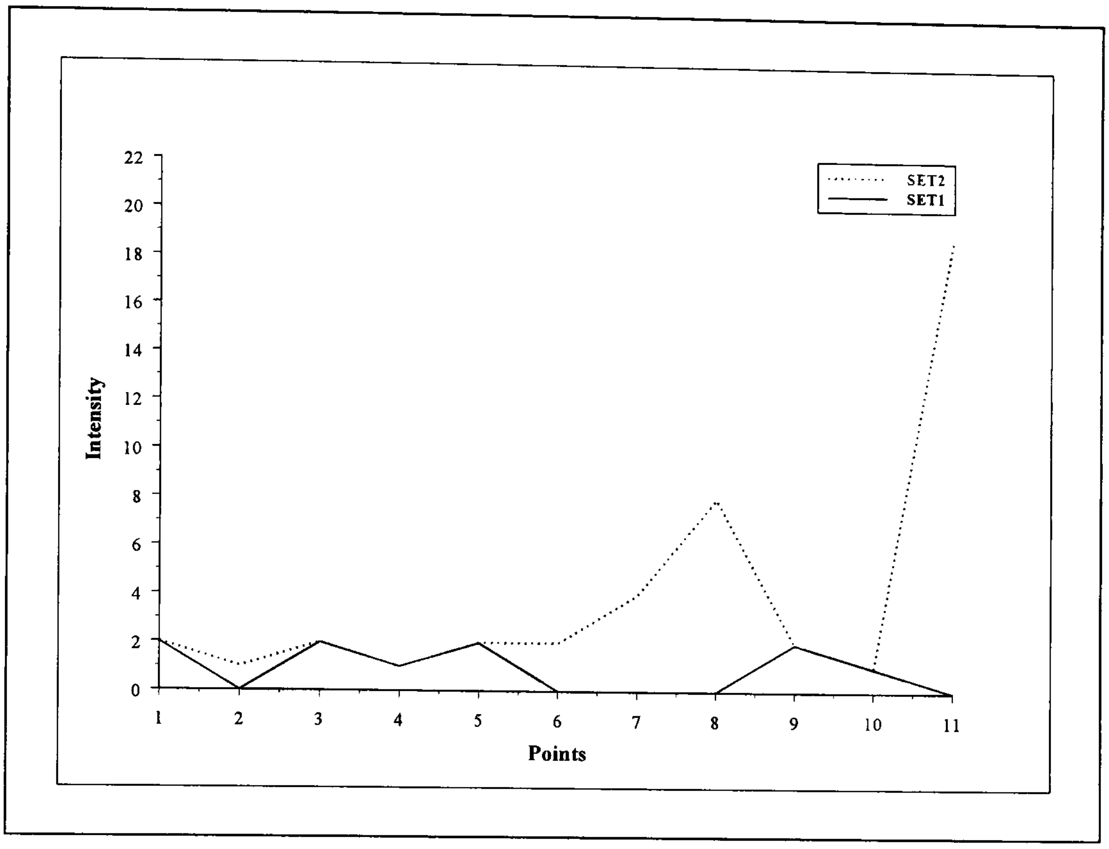

Figure 3.6: Data Sets - $I_1(f)/I_2(f)$

Figure 3.7 illustrates the discrete cross-correlation spectrum $R_{12}(\tau)$ obtained from the comparison of $I_1(f)$ and $I_2(f)$ employing Equation 3.5. The spectrum was constructed from the discrete values extracted from Equation 3.5 employing $\tau = 0$ before summation was applied.

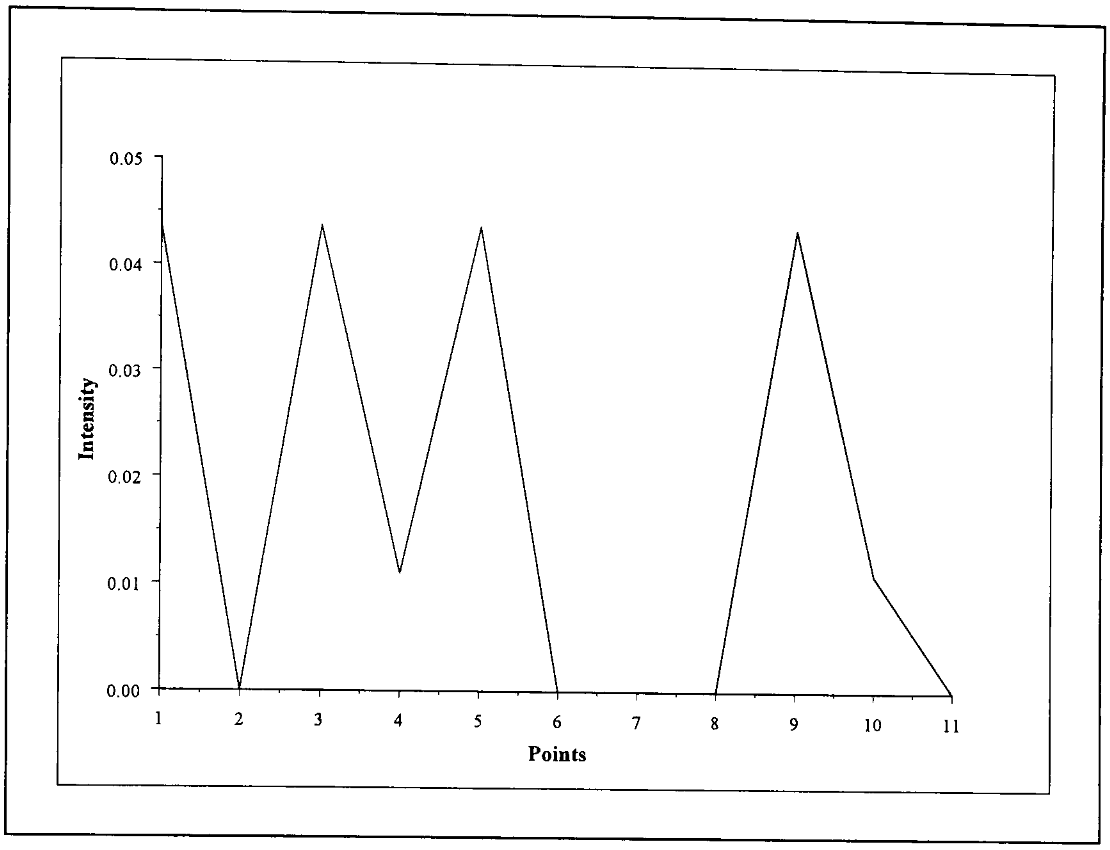

Figure 3.7: Discrete correlation analysis - $R_{12}(\tau)$

The discrete correlation analysis $R_{12}(\tau)$ of Figure 3.7 illustrates that large discrepancies are incurred at positions: 2; 6; 7; 8; and 11, indicated by zero values. However, the magnitude of these errors are not indicated, and rank ordering of these discrepancies is not achieved. It is therefore difficult to guide a user to areas of significant error within a comparison. The correlation coefficient R_{12} = 0.197 does however indicate that the comparison of $I_1(f)$ and $I_2(f)$ incurs major discrepancies.

Example 3.2: Discrete correlation analysis - $I_1(f)/I_3(f)$.

Figure 3.8 illustrates a comparison of $I_1(f)$ with a new signal $I_3(f)$, indicating a significant improvement over the comparison of Figure 3.6. Whilst Figure 3.9 illustrates the discrete correlation analysis described in Example 3.1, applied to the comparison of Figure 3.8.

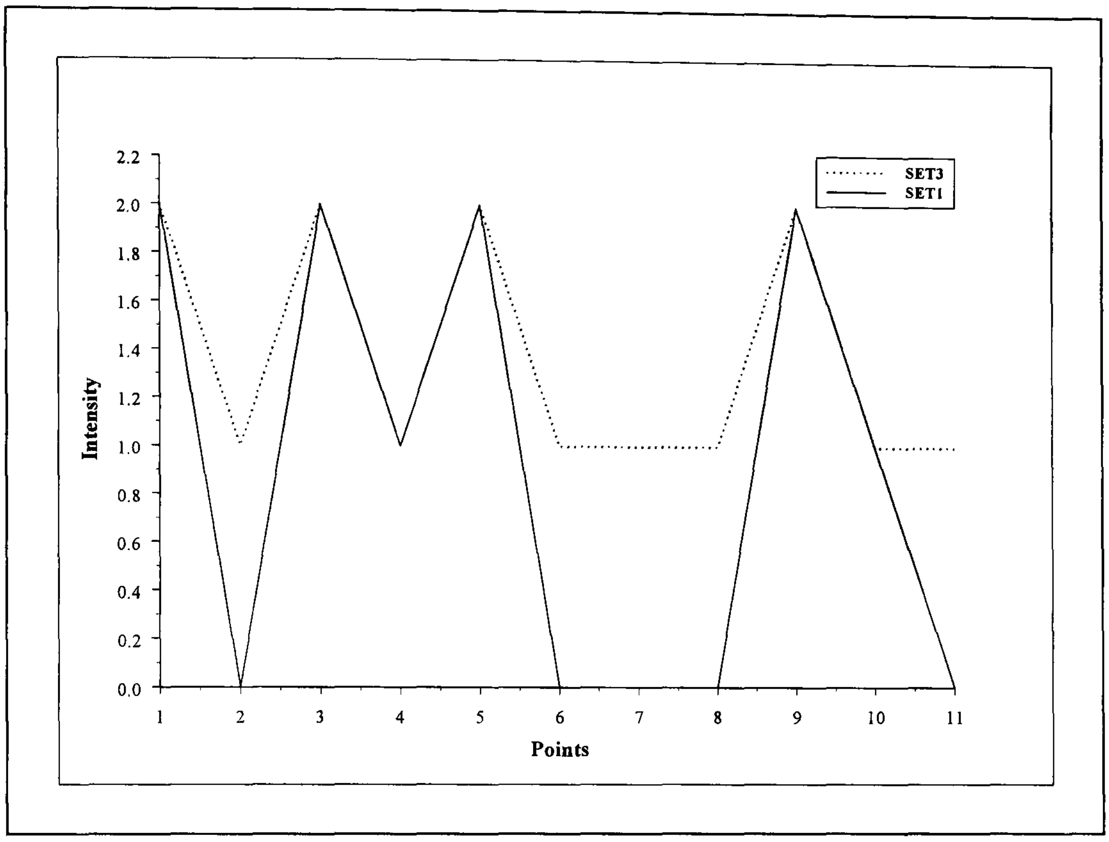

Figure 3.8: Data Sets - $I_1(f)/I_3(f)$

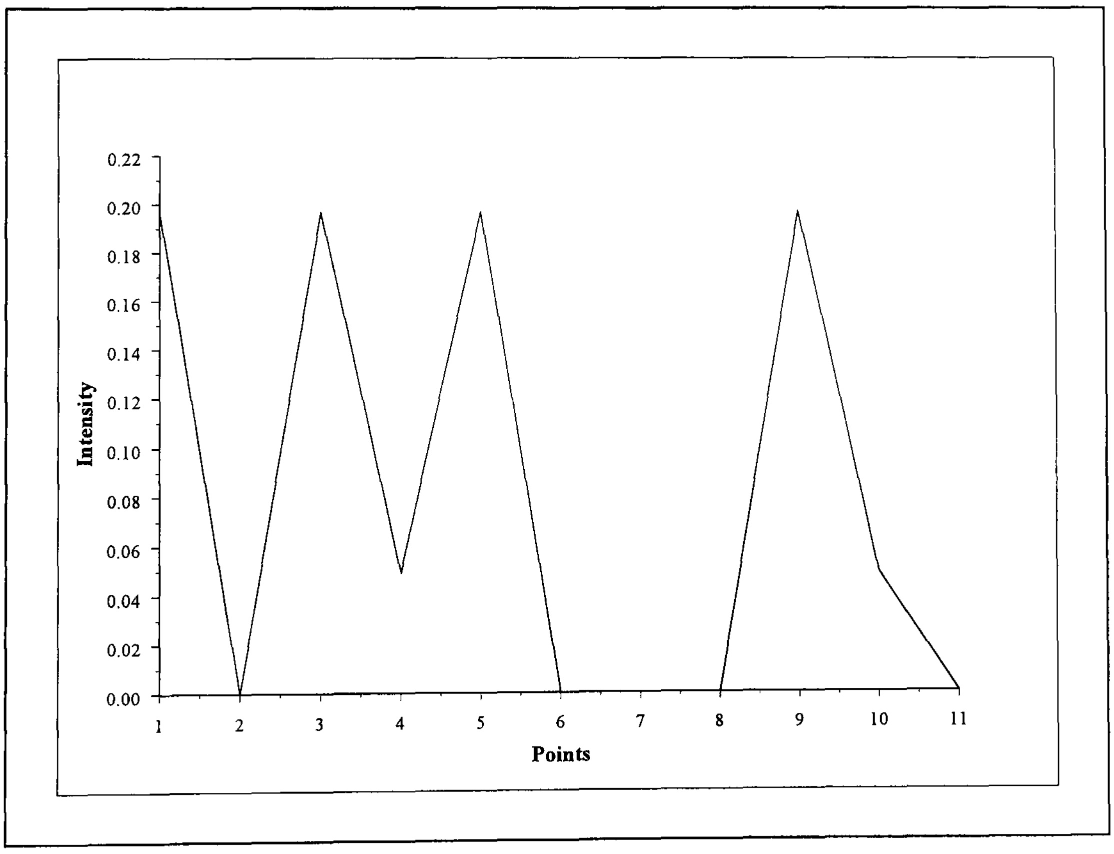

Figure 3.9: Discrete correlation analysis - $R_{13}(\tau)$

The results of $R_{13}(\tau)$ illustrated in Figure 3.9, indicate that discrepancies between the compared data signals of Figure 3.8 are located at 2, 6, 7, 8 and 11. However, the results of Figure 3.9 do not indicate significant reductions in the discrepancies at these locations, with respect to the results illustrated in Figure 3.7. The global figures of merit obtained from Examples 3.1 and 3.2, do however indicate that the comparison of Figure 3.8 is significantly better than that of Figure 3.6. This is indicated by the correlation coefficients $R_{12} = 0.197$ and $R_{13} = 0.885$.

3.1.4 Section Summary

Correlation techniques have proved to be useful tools in assessing the overall similarity between compared signals. However, weaknesses lie in their lack of discernment in providing diagnostic or discrete evaluations. Correlation methods may, as discussed, be split into several key measures indicating: cross-correlation; RMS symmetry; and RMS difference. Each of these measures provides information about the global similarity between compared signals however, these measures do not mirror the mechanisms employed by engineers and scientists undertaking visual evaluations, namely: atomic, relational and positional discrepancies/similarities outlined in Chapter 2. Furthermore, no specific qualitative scaling is associated with the method of correlation and although a great deal of information is extracted from an evaluation, it is cumbersome, not lending itself to a straightforward interpretation in terms of the categories and classifications employed by humans except in isolated application areas. A further weakness in the method of correlation is its inability to overcome problems due to the inherent characteristics of the compared signals or the application area under investigation. As evaluations are of a rigid nature, with a predefined structure which can not be altered by the user in terms of either amplitude levels or feature differences which comprise the main measures employed in a visual inspection of signals.

3.2 RELIABILITY FACTORS

Reliability factors move one step further in the search for accurate, meaningful and objective automated validation methodologies. Reliability factors were developed to analyse discrepancies between experimental and modelled results in the field of low energy electron diffraction, but are applicable to wider areas of study. They compare signals using simple difference measures, with emphasis placed on critical areas of the compared signals employing derivatives. The inherent nature of reliability factors allows for the potential application of discrete analyses, which if implemented provide informative diagnostic data.

3.2.1 Zanazzi Jona Reliability Factor

The first reliability factor introduced is that of Zanazzi and Jona which was devised in 1977. In developing the reliability factor, Zanazzi and Jona chose to emphasise the importance of matching peak positions as opposed to peak heights. This is implemented in a comparison of differing slopes, employing first derivatives of compared intensities with respect to energy, over the full spectrum of the compared signals, such that:

$$F(f) = |I'_{SET1}(f) - cI'_{SET2}(f)| \qquad \textbf{(3.8)}$$

where single primes denote first derivatives.

Equation 3.8 is essentially a gradient emphasis and difference algorithm. First derivatives of the compared intensities are evaluated in an attempt to emphasise the shapes and positions of peaks/troughs embedded in the full spectra of the signals under investigation. Whilst second derivatives are introduced to emphasise differences between narrow features, in an attempt to emphasis the shapes and more importantly the positions of peaks/troughs through the use of a weighting factor *W(f)*, such that:

$$W(f) = \frac{|I''_{SET1}(f) - cI''_{SET2}(f)|}{|I'_{SET1}(f)| + \varepsilon} \qquad \textbf{(3.9)}$$

where

$$\varepsilon = |I'_{SET1}(f)|_{\max} \tag{3.10}$$

where single and double primes denote first and second derivatives respectfully.

Furthermore, the first and second derivatives employed in Equations 3.8 and 3.9 are normalised to the average intensity c of the compared signals, given by:

$$c = \frac{\sum_{f_{\min}}^{f_{\max}} I_{SET1}(f)}{\sum_{f_{\min}}^{f_{\max}} I_{SET2}(f)} \tag{3.11}$$

The normalisation factor c has the effect of suppressing the importance of the relative intensity difference between compared signals, allowing assessments to be made based on the shapes and positions of features whilst suppressing the importance of amplitude differences.

The final integral R_{ZJ} is given in Equation 3.13, with A or the reducing factor introduced to eliminate the dependence of the Reliability Factor on the intensity of the reference signal. The reducing factor A is given in equation 3.12:

$$A = \frac{\delta e}{\sum_{f_{\min}}^{f_{\max}} I_{SET1}(f)} \tag{3.12}$$

$$R_{ZJ} = \frac{A}{\delta e} \sum_{f_{\min}}^{f_{\max}} W(f)F(f) \tag{3.13}$$

where

$$\delta e = f_{\max} - f_{\min}$$

Employing the analysis described in Equations 3.8 - 3.13, a Zanazzi Jona reliability value R_{ZJ} of 0.107 was obtained for the comparison illustrated Figure 3.10.

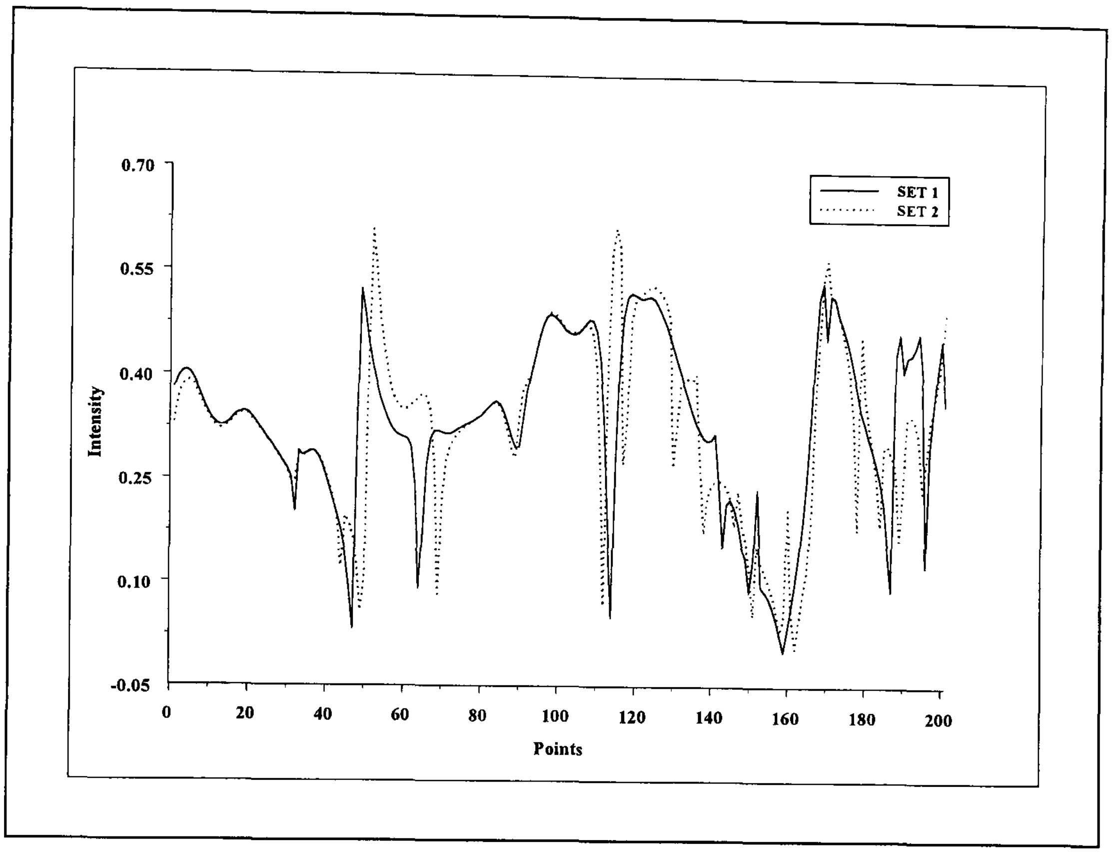

Figure 3.10: Data Sets - $I_{SET1}(f)/I_{SET2}(f)$

3.2.1.1 Discussion

The reliability value R_{ZJ} indicates a global figure of merit based on the quality of a comparison between two complex signals. This value comprises first and second order derivative differences indicating relational and positional discrepancies, but does not take into account the atomic (amplitude) differences between compared signals. In assessing the quality of a comparison, the method of Zanazzi Jona provides a rigorous analysis of the relative discrepancies between the shapes and positions of features inherent in the signals under investigation. However, this type of analysis lacks discernment in cases where the relative amplitude levels are of great importance in obtaining a true evaluation of the overall differences between signals. Furthermore, the complexity of the underlying method developed by Zanazzi and Jona complicates the process of modifying the algorithms derived in Equations 3.8 - 3.13. In attempting to revise the method, Equation 3.13 must primarily be split into two separate measures of discrepancy. With the two measures providing separate information on first and second order derivative differences (relational discrepancies) between compared signals. A further measure is also required to assess differences between compared signal amplitude levels (atomic discrepancies), allowing a complete validation methodology to be obtained. In order to provide information expressing specific areas of discrepancy between two signals, separate measures dedicated to analysing isolated areas of the compared signals must be developed. Sections 3.2.2 and 3.2.3 along with Chapter 4 detail advances in the area of quantitative validation employing separate measurements which provide information expressing differences between compared amplitude levels and feature shapes and positions.

3.2.2 Van Hove Reliability Factor - Van Hove I

The second example of reliability factors introduced in this Chapter, is that of M. A. van Hove *et al* 1997, who define a five formulae reliability factor. Van Hove identified that comparisons of data sets most commonly employ measures of position and width of peaks; shape of peaks, shoulders and valleys; number of peaks and shoulders and their relative heights. The five formulae chosen by van Hove are summarised below:

$$R_1 = \frac{\sum_{f_{\min}}^{f_{\max}} \left| I_{SET1}(f) - cI_{SET2}(f) \right|}{\sum_{f_{\min}}^{f_{\max}} \left| I_{SET1}(f) \right|} \qquad \textbf{(3.14)}$$

$$R_2 = \frac{\sum_{f_{\min}}^{f_{\max}} \left(I_{SET1}(f) - cI_{SET2}(f) \right)^2}{\sum_{f_{\min}}^{f_{\max}} \left(I_{SET1}(f) \right)^2} \qquad \textbf{(3.15)}$$

where f denotes an energy point on the response curve being evaluated and c (described in section 3.2.1) is a scaling factor equated to the ratio between the average intensities in the reference and comparison curves, such that:

$$c = \frac{\sum_{f_{min}}^{f_{max}} I_{SET1}(f)}{\sum_{f_{min}}^{f_{max}} I_{SET2}(f)} \quad \textbf{(3.16)}$$

Van Hove reported that both R_1 and R_2 tend to emphasise the match in positions, heights and widths of peaks and valleys. However they pay little attention to the number of shoulders and bumps which describe a single peak. R_1 and R_2 would also fail in distinguishing curvatures of peaks; e.g. a peak having a predominantly concave or convex curvature. To eliminate the shortcomings of R_1 and R_2, van Hove proposed three new R-factors, given below:

$$R_3 = \text{Fraction of energy range with slopes of different signs} \quad \textbf{(3.17)}$$

where

$$R_3 = \frac{Number\ \ of\ \ Positive\ \ Samples\ \ in\ \ I'_{SET1}(f)}{Number\ \ of\ \ Positive\ \ Samples\ \ in\ \ I'_{SET2}(f)}$$

$$R_4 = \frac{\sum_{f_{min}}^{f_{max}} \left| I'_{SET1}(f) - cI'_{SET2}(f) \right|}{\sum_{f_{min}}^{f_{max}} \left| I'_{SET1}(f) \right|} \quad \textbf{(3.18)}$$

$$R_5 = \frac{\sum_{f_{min}}^{f_{max}} \left(I'_{SET1}(f) - cI'_{SET2}(f) \right)^2}{\sum_{f_{min}}^{f_{max}} \left(I'_{SET1}(f) \right)^2} \quad \textbf{(3.19)}$$

where single primes denote first derivatives of intensity with respect to energy. R_3 selects the portions of a feature that are of differing slope, while R_4 and R_5 are sensitive to the slopes of the data sets.

Table 3.2 indicates the results of the van Hove reliability factor employing the comparison illustrated in Figure 3.11. Where R_T - derived by the author - denotes an evaluation of the total discrepancy between compared signals and is equated to the vector addition of the five reliability factors previously defined, such that:

$$R_T = \sqrt{R_1^2 + R_2^2 + R_3^2 + R_4^2 + R_5^2} \tag{3.20}$$

Table 3.2 indicates the values derived for each of the five formulae of the van Hove reliability factor, along with the overall or total reliability R_T of the comparison derived in Equation 3.20.

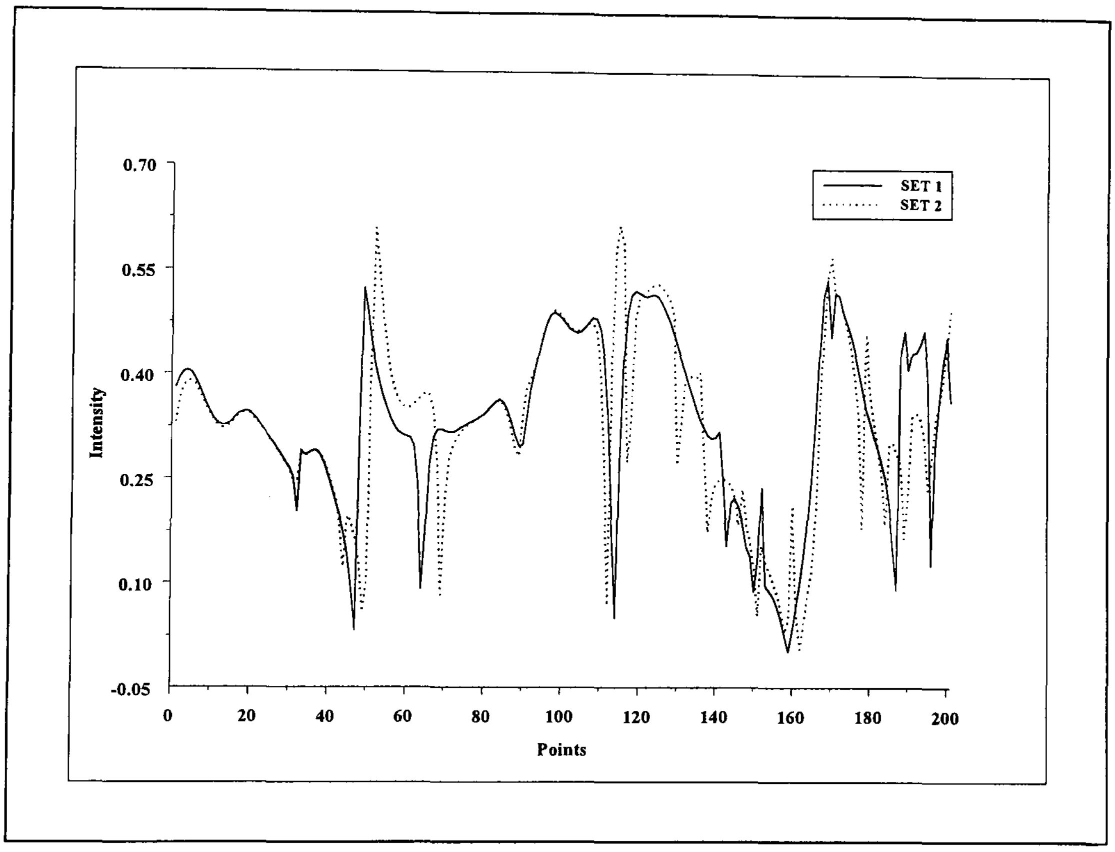

Figure 3.11: Data Sets - $I_{SET1}(f)/I_{SET2}(f)$

R_1	R_2	R_3	R_4	R_5	R_T
0.161	*0.071*	*0.258*	*1.443*	*2.57*	*2.96*

Table 3.2: Van Hove reliability factor results - $I_{SET1}(f)/I_{SET2}(f)$

3.2.2.1 Discussion

The reliability values R_1, R_2, R_3, R_4, R_5 and R_T included in the method of van Hove, introduce the concept of a multilevel validation scheme. The five reliability values evaluate areas of discrepancy between compared data signals, allowing assessments to be broken down into isolated measurements. R_1 and R_2 indicate discrepancies between amplitude levels or atomic differences, whilst R_3, R_4 and R_5 denote discrepancies between the shapes and positions of features embedded in the compared signals. R_T evaluates the total difference between compared signals in terms of both amplitude levels and feature shapes and positions. Furthermore, whilst an instantaneous or point by point evaluation of R_1, R_2, R_4, and R_5 is possible, van Hove did not include this in-depth diagnostic analysis in the method detailed in Equations 3.14 - 3.20. This failure to incorporate diagnostic analyses invariably weakens the van Hove reliability factor in cases where high levels of diagnostic information are required for feedback.

3.2.3 Proposed Modification To Van Hove Reliability Factor – Van Hove II

Modifications - by the author AJM Martin - to the original reliability factors R_1, R_2, R_4 and R_5 of van Hove *et al* originally derived in Equations 3.14, 3.15, 3.18 and 3.19 respectively, allow both figures of merit and instantaneous reliability responses curves to be realised for R_1, R_2, R_4 and R_5. These modifications increase the methods ability to produce in-depth diagnostic information concerning the quality of a comparison. The nature of the reliability factor R_3 derived in Equation 3.17 however, does not lend itself to such an analysis as the general form of the equation relies on information based on the full spectra of the compared signals. The four modified reliability factors are summarised in Equations 3.21 - 3.32.

$$R_1 = \frac{1}{\left(f_{\max} - f_{\min}\right)} \sum_{f_{\min}}^{f_{\max}} R_1(f) \qquad \textbf{(3.21)}$$

$$R_1(f) = \frac{\left| I_{SET1}(f) - cI_{SET2}(f) \right|}{N_1} \qquad \textbf{(3.22)}$$

$$N_1 = \frac{0.25}{\left(f_{\max} - f_{\min}\right)} \sum_{f_{\min}}^{f_{\max}} \left| I_{SET1}(f) \right| \tag{3.23}$$

$$R_2 = \frac{1}{\left(f_{\max} - f_{\min}\right)} \sum_{f_{\min}}^{f_{\max}} R_2(f) \tag{3.24}$$

$$R_2(f) = \frac{\left(I_{SET1}(f) - cI_{SET2}(f)\right)^2}{N_2} \tag{3.25}$$

$$N_2 = \frac{1}{\left(f_{\max} - f_{\min}\right)} \sum_{f_{\min}}^{f_{\max}} \left(I_{SET1}(f)\right)^2 \tag{3.26}$$

$$R_4 = \frac{1}{\left(f_{\max} - f_{\min}\right)} \sum_{f_{\min}}^{f_{\max}} R_4(f) \tag{3.27}$$

$$R_4(f) = \frac{\left| I'_{SET1}(f) - cI'_{SET2}(f) \right|}{N_4} \tag{3.28}$$

$$N_4 = \frac{0.167}{\left(f_{\max} - f_{\min}\right)} \sum_{f_{\min}}^{f_{\max}} \left| I'_{SET1}(f) \right| \tag{3.29}$$

$$R_5 = \frac{1}{\left(f_{\max} - f_{\min}\right)} \sum_{f_{\min}}^{f_{\max}} R_5(f) \tag{3.30}$$

$$R_5(f) = \frac{\left(I'_{SET1}(f) - cI'_{SET2}(f)\right)^2}{N_5} \tag{3.31}$$

$$N_5 = \frac{0.167}{\left(f_{\max} - f_{\min}\right)} \sum_{f_{\min}}^{f_{\max}} \left(I'_{SET1}(f)\right)^2 \qquad \textbf{(3.32)}$$

where N_1, N_2, N_4 and N_5 denote normalisation factors derived empirically, applied to R_1, R_2, R_4 and R_5 respectively. The four normalisation factors are applied in an attempt to gain a consistent scaling methodology between the instantaneous reliability factor response curve values and the global reliability measures extracted from the method. Furthermore, within this scaled methodology a revised reliability factor R_3 is equated to the difference between the ratio of samples of differing signs for the signals under investigation, given by:

$$R_3 = Ratio_1 - Ratio_2 \qquad \textbf{(3.33)}$$

where

$$Ratio_1 = \frac{Number\ \ of\ \ Positive\ \ Samples\ \ in\ \ I'_{SET1}(f)}{Number\ \ of\ \ Negative\ \ Samples\ \ in\ \ I'_{SET1}(f)} \qquad \textbf{(3.34)}$$

$$Ratio_2 = \frac{Number\ \ of\ \ Positive\ \ Samples\ \ in\ \ I'_{SET2}(f)}{Number\ \ of\ \ Negative\ \ Samples\ \ in\ \ I'_{SET2}(f)} \qquad \textbf{(3.35)}$$

Table 3.3 indicates the results of the revised reliability factors of van Hove, employing the comparison illustrate in Figure 3.12, Equations 3.21 - 3.35 and the total reliability factor derived in Equation 3.20. Whilst Figures 3.13 - 3.16 illustrate the van Hove reliability response curves $R_1(f)$, $R_2(f)$, $R_4(f)$ and $R_5(f)$ respectively.

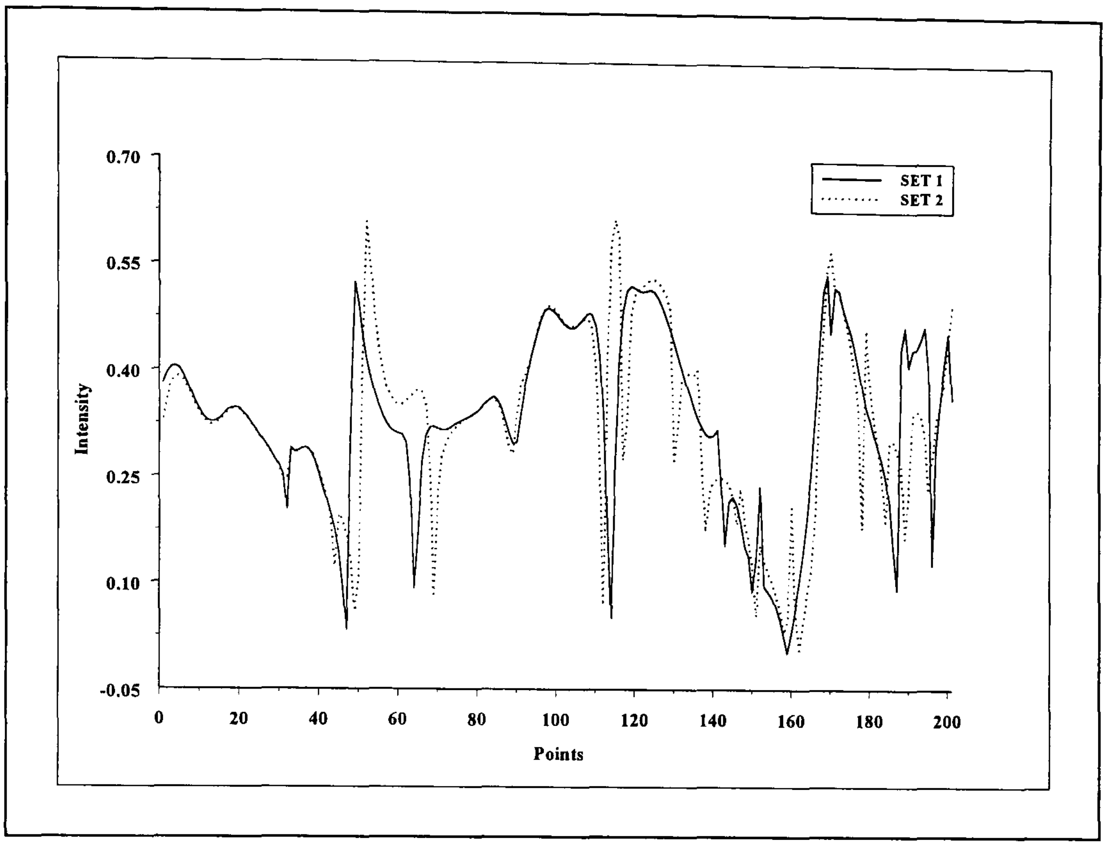

Figure 3.12: Data Sets - $I_{SET1}(f)/I_{SET2}(f)$

R_1	R_2	R_3	R_4	R_5	R_T
0.0402	*0.0711*	*0.3866*	*0.2404*	*0.4276*	*0.6298*

Table 3.3: Modified van Hove reliability factor results - $I_{SET1}(f)/I_{SET2}(f)$

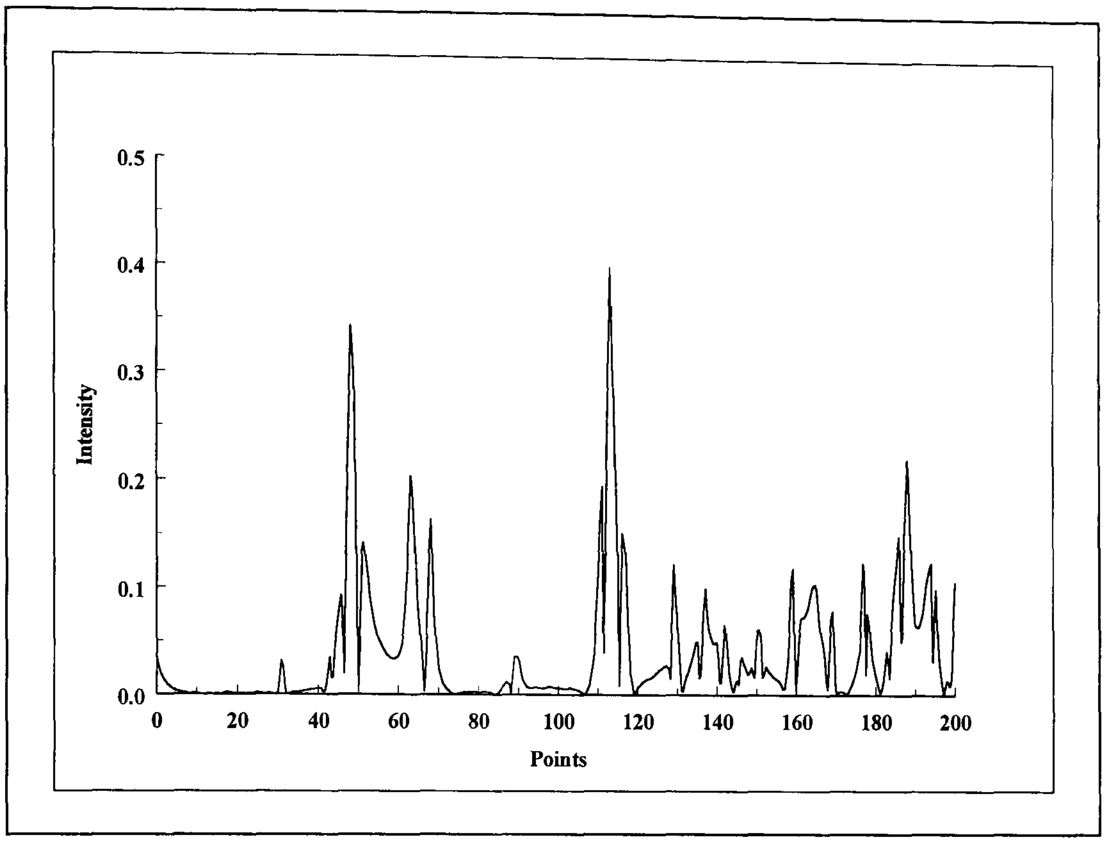

Figure 3.13: Modified van Hove reliability factor - $R_1(f)$

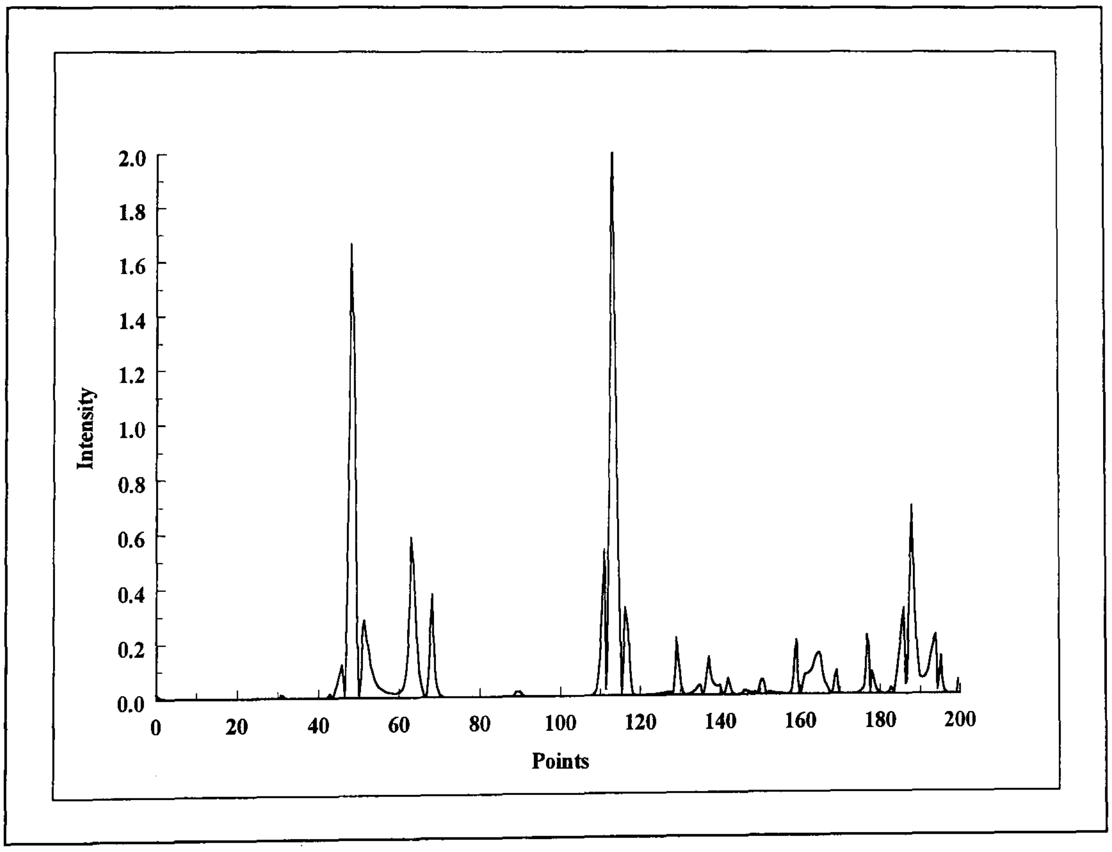

Figure 3.14: Modified van Hove reliability factor - $R_2(f)$

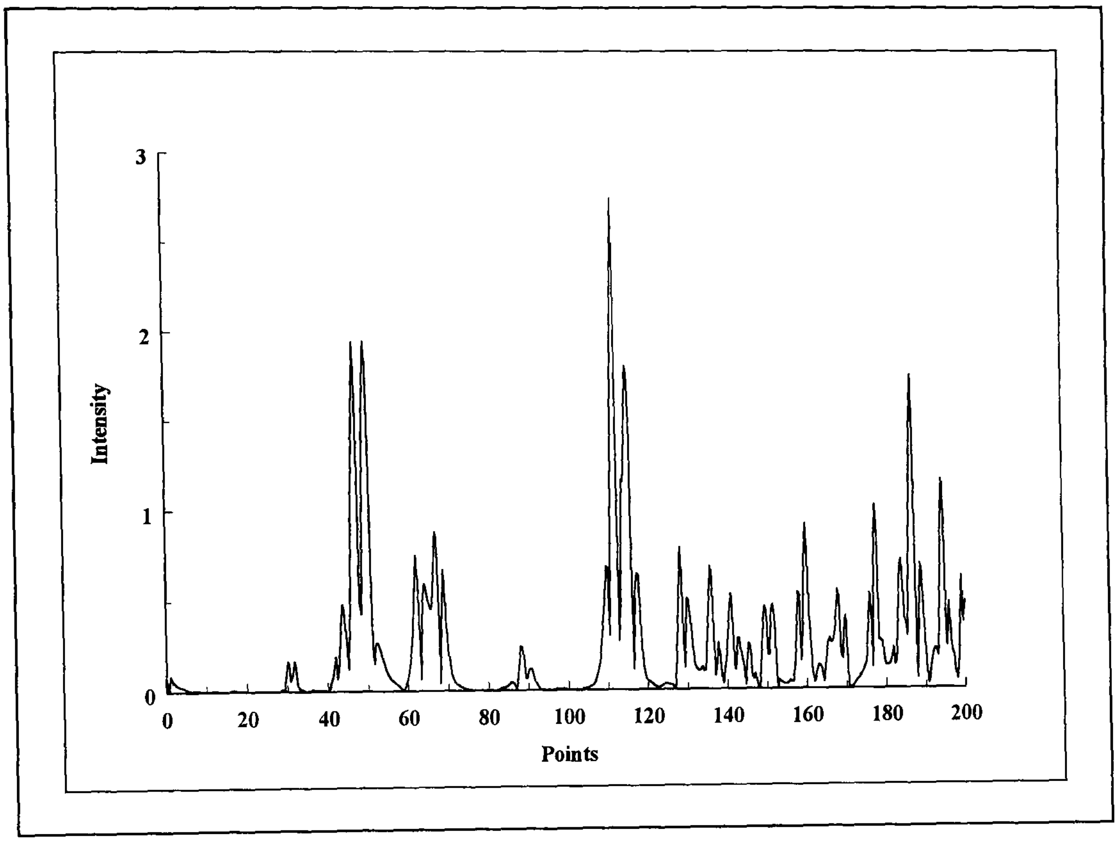

Figure 3.15: Modified van Hove reliability factor - $R_4(f)$

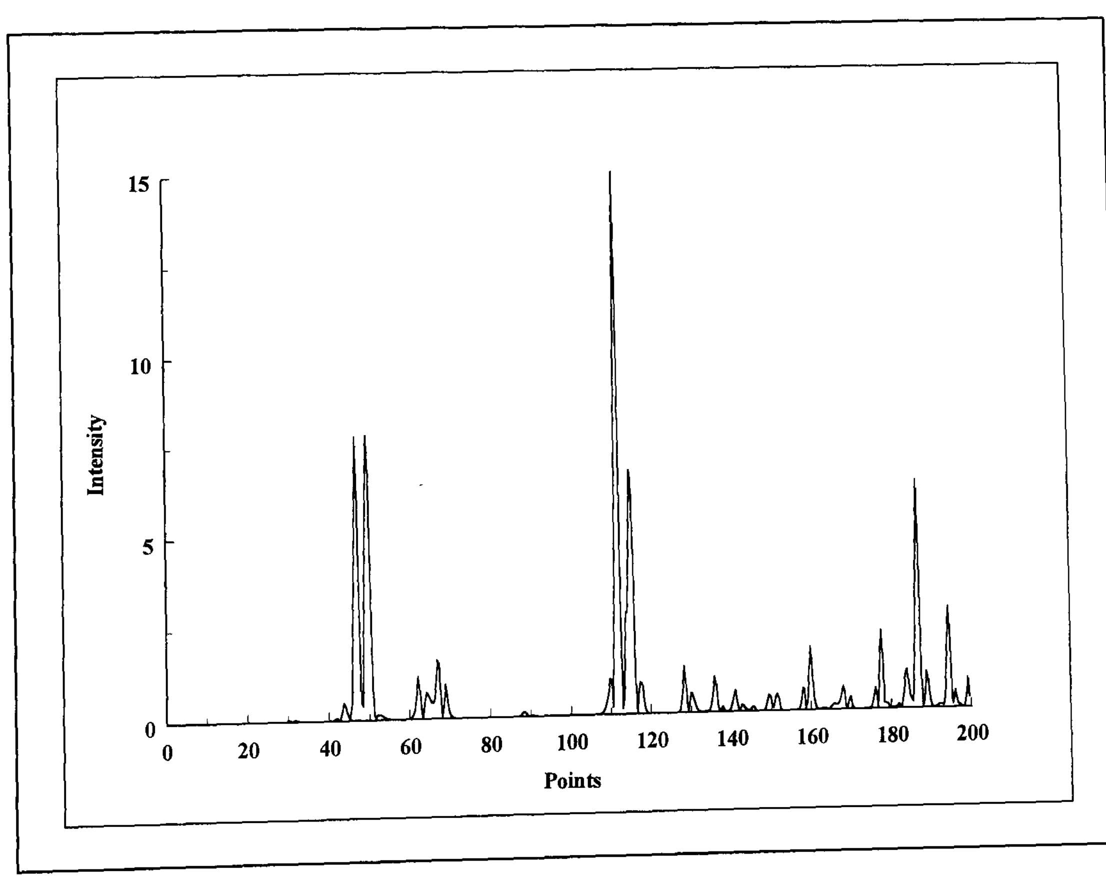

Figure 3.16: Modified van Hove reliability factor - $R_5(f)$

3.2.3.1 Discussion

The modified van Hove reliability factors derived in Equations 3.21 - 3.33 provide a significant amount of information on the nature, location and magnitude of discrepancies between compared signals. However, the reliability factors: R_1 and R_2; and R_4 and R_5 may be viewed as identical measurement pairs with different scaling. R_1 and R_2 both evaluate discrepancies between amplitude levels, whilst R_4 and R_5 evaluate discrepancies between the shapes and positions of features embedded in the compared signals. Where R_2 and R_5 are amplified versions of R_1 and R_4 respectively, illustrated in Figures 3.14 - 3.17. In order to further the method of van Hove, R_2 may be discarded and R_1 scaled to compensate for the loss of this measure. Whilst R_5 may be replaced with an evaluation of second order derivative differences, thus emphasising discrepancies between the shapes and positions of trends and features within the signals under investigation. These steps would allow an analysis employing four inherently different algorithms dedicated to evaluating discrepancies between complex data signals. This theory is pursued further in Chapter 4.

3.2.4 Section Summary

The results from both Zanazzi Jona and van Hove show the possible benefits achievable through the use of formal validation methods. In the case of van Hove the incorporation of individual measurements has given an insight into the potential benefits of separating validation information into specific subsets of discrepancies between compared data signals. This is one step towards an ideal analysis of compared signals and is pursued further in Chapter 4.

The advantage of the Zanazzi Jona reliability factor is that both first and second derivatives are evaluated, and their differences assessed, providing highly sensitive information on both the positions and shapes of compared features. The advantage of the van Hove reliability factor is that whilst still sensitive to peak positions, only an evaluation of first derivatives is necessary. However, in both the method of Zanazzi Jona and van Hove, the results obtained from one comparison are inflexible and isolated compared to those gained from another. This isolation and inflexibility is due to the algorithms inconsistent scaling when confronted with several sets of comparisons with a broad range of different characteristics from widely different application areas (detailed in Chapter 5).

The real benefit of using either of the reliability factor methods presented here would be noticed within a single application area where the relative differences between the inherent characteristics of the results are less marked. It is to this end that reliability factors provide useful information. Two such factors, namely Zanazzi Jona and van Hove, were implemented with very encouraging results.

3.3 SUMMARY OF TECHNIQUES

Correlation in some part allows a quantitative view of the discrepancies between two sets of data, but lacks both the consistency and the vital information content necessary to allow a valid judgement on the nature and quality of a comparison. In order to confirm that, for example, a numerical modelling technique adequately simulates the behaviour of a real system (experiment), historically, validation has been undertaken on a 'case study' basis. Repeated comparisons are made and these gradually inspire confidence through competent application. In order for technological systems to be used with confidence, a high correlation between comparison data sets must be made. However, within application areas such as EMC or r.f. engineering, correlation techniques are not in wide spread use due to the complexity of the signals under investigation. Generally, 'global' techniques such as correlation and reliability factors are not powerful enough to pick out the subtleties which an experienced technologist will naturally gravitate towards. To date, this reason has limited the ability to automate assessments of complex comparisons.

Multiplication (similarity) methods such as correlation algorithms can not provide in-depth information due to the nature of the analysis applied to compared signals. Within the method of correlation, an enormous amount of analysis information is locked into the normalisation factor. That is to say, that when the point by point correlation results of a single shift are extracted, they do not fully realise the errors within the system at the instantaneous points analysed. This is not to say that the method of correlation does not work, but merely that the normalisation factor employed holds specific information regarding the discrete validation results, which can not be fully unlocked to provide instantaneous results.

Difference algorithms, such as those employed in reliability factors are not subject to this problem, allowing results to be extracted on a point by point basis. These types of analyses also allow for global figures of merit to be equated and are potentially suitable for methods of higher analyses to be applied, such as statistics. One such method detailed in Section 3.2.3, provides a significant amount of information concerning the nature, location and magnitude of discrepancies between compared signals. However the measures employed to provide this information do not accurately mirror the method of visual evaluation. This in turn adversely affects the methods ability to rank order comparisons in the same way as the combined interpretations of subjects performing identical tasks.

3.4 CHAPTER SUMMARY

Whilst the human brain is the most powerful recognition device known, the level of attention, focus and acuity of subjects are all limited when several comparisons of complex data are to be validated. Visual evaluation is the most prevalent form of data analysis known to date, however, experiential and physical differences between subjects performing visual evaluation tasks along with age, regularly cause variabilities between assessment results. Far from regarding this as a problem, the variability between interpretations made by experienced technologists is a real phenomenon underlining the complexity of the original data and should be something which automated schemes must reflect. In order to make further progress in the field of automated validation, there is a clear need to carefully study the process by which humans inspect and compare data sets, and methods in which this capability may be transferred to machines.

Inevitably, there still remain tasks which can not be automated within complicated validation routines, and it is good practice to divide validation schemes between man and machine, allowing humans to engage the tasks which can not easily be replicated by computers. Hence, the goal of an automated validation scheme should be to mirror visual evaluations undertaken by highly skilled engineers within their area of expertise, with the information gained from an analysis being presented in a categorical manner which is directly related to 'human' interpretations.

It is proposed that the fundamental requirements of any successful automated validation technique are covered by the six points below:

1. ***Implementation of the validation technique must be simple.***
2. ***The technique must be computationally straightforward, requiring at most a modest processor speed and relatively little memory, while retaining the ability to produce validation results rapidly.***
3. ***The method must possess the ability to mirror human perceptions, based on criteria taken into account by the brain during a visual comparison of two data sets.***
4. ***The method must be flexible, with the ability to validate data taken from a wide cross section of areas.***
5. ***Validation results obtained must be simple to interpret, with little or no training required.***
6. ***The technique must produce levels of information from a single pass/fail value to a full - point by point - diagnosis of the compared signals.***

In theory, the basic numerical tools (filter algorithms, differential calculus and statistics) exist to automate visual evaluations of complex data signals, however, there still remain key aspects of validation procedures which may not to date be automated. In retaining partial human interaction within modern validation schemes however, a level of flexibility and subjectivity may, if necessary, be incorporated into otherwise rigid assessments of potentially complex data signals. It is this flexibility along with a universal scaling methodology that will permit data signals to be analysed from a wide cross section of application areas allowing the realisation of true multidisciplinary validation schemes.

CHAPTER 4

FEATURE SELECTIVE VALIDATION

4. FEATURE SELECTIVE VALIDATION

Throughout a visual evaluation of complex data signals, the brain must sort an endless stream of data received from the eyes into meaningful information. From the chaotic jumble of data provided by the visual system, the brain must decide what parts are meaningful and what parts are trivial in order to efficiently produce a coherent picture of the stimulus under investigation. This phenomenon is known as perception. Human perception is the best recognition system known to date, and while this process may be slow and expensive the combined results of visual evaluations undertaken by skilled technologists possess high levels of confidence. A further component of perception is the brains overriding tendency, whether consciously or unconsciously, to categorise stimuli, giving each a name. It is common to see the terms "good" or "excellent" used to describe the quality of comparisons in published work. However, the definitions of these terms are entirely subjective and, it may be conjectured, that these terms are used without quantitative assessments to reinforce them.

The fundamental problem remaining in the field of validation is that of rapidly assessing and differentiating the level of agreement or disagreement and thus quantifying comparisons between potentially complex waveforms. Assessments must be accurate and robust, producing repeatable validation information which mirrors the information extracted by engineers or scientists undertaking visual evaluations. However, if the aim of an automated validation scheme is to provide a reliable mirror of the experienced technologist, it is important to understand that the behaviour of the system depends fundamentally on the extent to which an engineer responds to the information obtained from a comparison. Such dynamic behaviour is difficult to predict, and the design of quantitative validation procedures to achieve acceptable response is not a trivial matter. Information extracted from an automated validation scheme expressing the quality of a comparison must employ conventional interpretations. These interpretations are conventions created by humans in order to communicate in a meaningful way. Furthermore, these validation techniques must be flexible, allowing valid assessments of comparison signals from a wide variety of application areas.

Correlation and Reliability factors are automated validation methods in widespread use to quantify the level of agreement or dissimilarity between compared data sets, however, these current automated validation procedures encounter many problems. In order to overcome these problems, automated validation methodologies must be approached from a new stance. Not attempting to produce an entirely different process to that of visual evaluation, but simulating the existing technique, whilst enhancing the method, and producing several levels of recordable and repeatable information. Reasons for automating visual evaluations include: the need to control variabilities between visual assessment results; the reduction of cost (a skilled engineer is an expensive commodity); the desire to reduce ambiguities caused by fatigue; and the inability of humans to process and cache extremely large volumes of data.

The methods of both correlation and Zanazzi Jona detailed in Chapter 3 employ single validation algorithms which derive values that quantify the overall differences between two data signals. This approach is adequate if an assessment of the overall differences between signals is required. However, in cases where precise feedback is required, it is more appropriate to develop a number of individual algorithms[van Hove 1997] to assess the quality of a comparison. Furthermore, these individual measures must produce information directly related to the combined results of skilled engineers undertaking visual evaluations.

For a validation technique to be successful, information must be provided on the location and magnitude of major discrepancies between compared signals. That is to say, a discrete analysis must be provided, along with figures of merit expressing the total discrepancy between compared data signals. Without this information, a full investigation of a comparison can not be accomplished. Furthermore, for a validation technique to be embraced by those who use it, a full understanding of the manner in which it operates must first be realised.

This Chapter augments the techniques of correlation and reliability factors by detailing advances in a further automated validation method, the ***Feature Selective Validation*** (***FSV***) method. The FSV method was conceived as a technique to quantify the validation of numerical models by mirroring user perceptions. Specific novelty in this method is realised by the quantity of feedback information available to the user. Whilst the quantification of terms, such as 'a good comparison' can now be regarded as having a specific meaning. Furthermore, the FSV method allows automated comparisons of large volumes of complex data whilst reliably categorising the results into a common set of quality bands.

The FSV method comprises two component measures based on *amplitude differences* and *feature differences*. These measures are combined to form an overall assessment of the comparison in question or *global difference*. The three measures within the FSV method are strengthened by statistical analyses in the form of amplitude, feature and global confidence levels. Highly detailed diagnostic information on the location and magnitude of discrepancies is also made available through the employment of graphical (discrete) representations of the three measures.

The FSV method benefits from the ability to mirror human perception, whilst producing information which directly relates human variability and the confidence associated with it. The FSV method also builds on the common language of engineers and scientists alike, employing categories which relate to human interpretations of comparisons, namely: 'ideal', 'excellent', 'very good', 'good', 'fair', 'poor' and 'extremely poor'.

This chapter will consider the process requirements of successful automated validation schemes whilst detailing the development of the FSV method. The results presented illustrate the operation of the validation scheme, whilst Chapter 5 verifies its performance through the comparison of FSV information and the results of a validation trial involving a number of experienced engineers and technologists performing visual evaluations.

4.1 THEORY

The majority of complex data signals contain combinations of broad peaks/troughs (trends) and narrow peaks/troughs (features) on a base of amplitude levels. A quantitative comparison of results should, as a minimum, possess the ability to compare these fundamental features. In general, automated validation/verification methods require a multitude of distinct processing stages to accomplish true assessments of compared complex data signals. These include: the isolation of critical features within compared signals; the extraction and rank ordering of critical features; an evaluation of differences between extracted features; and the interpretation of information obtained from the validation scheme.

Automated validation schemes which support application areas of widely differing characteristics must be flexible, whilst retaining a rigid assessment of the differences between compared signals. It is hypothesised, that it is possible to develop a method of automated data validation utilising several processing and analysis stages, whilst employing a measured level of human input. These processing stages in turn must yield not only the correct interpretation of the data in question, but also place the correct emphasis on critical areas of the data signals. Furthermore, the information extracted from a comparison of data signals must relate to the confidence inherent in the combined results of subjects performing visual evaluation tasks.

The human visual search system may be viewed as the process of extracting critical features from stimuli, in terms of atomic, relational and positional extraction. Within the field of automated data signal validation, these three main extraction routines may be implemented through the employment of absolute values, and first and second order derivatives. Where first and second order derivatives are employed to obtain signatures representing the shapes and positions of both trends and features inherent in the signals under investigation. Employing this analysis, signature response curves may be constructed which emphasise both the shapes and positions of broad peaks/troughs (trends) and narrow peaks/troughs (features) which mirror the information employed by humans visually analysing complex signal comparisons. Figure 4.1 illustrates a comparison of complex data signals, whilst Figure 4.2 illustrates an enlarged version of the subsection indicated in Figure 4.1 denoting three data points related to one of the compared signals. Within Figure 4.2 the atomic parts of the signal are represented by the magnitudes (absolute values) of the three data points a, b and c. Whilst first and second order relational characteristics, namely: ab and bc, may be evaluated employing first and second order derivatives. Furthermore, as the analysis is undertaken on a discrete basis employing indexed samples, three levels of positional analysis may be obtained during an assessment. Where the indexed absolute values within an assessment allow an absolute analysis between the positions of amplitude levels. The indexed first derivative values allow a low level analysis between the positions of low order relational characteristics. Whilst indexed second derivative values allow a high level analysis between the positions of high order relational characteristics.

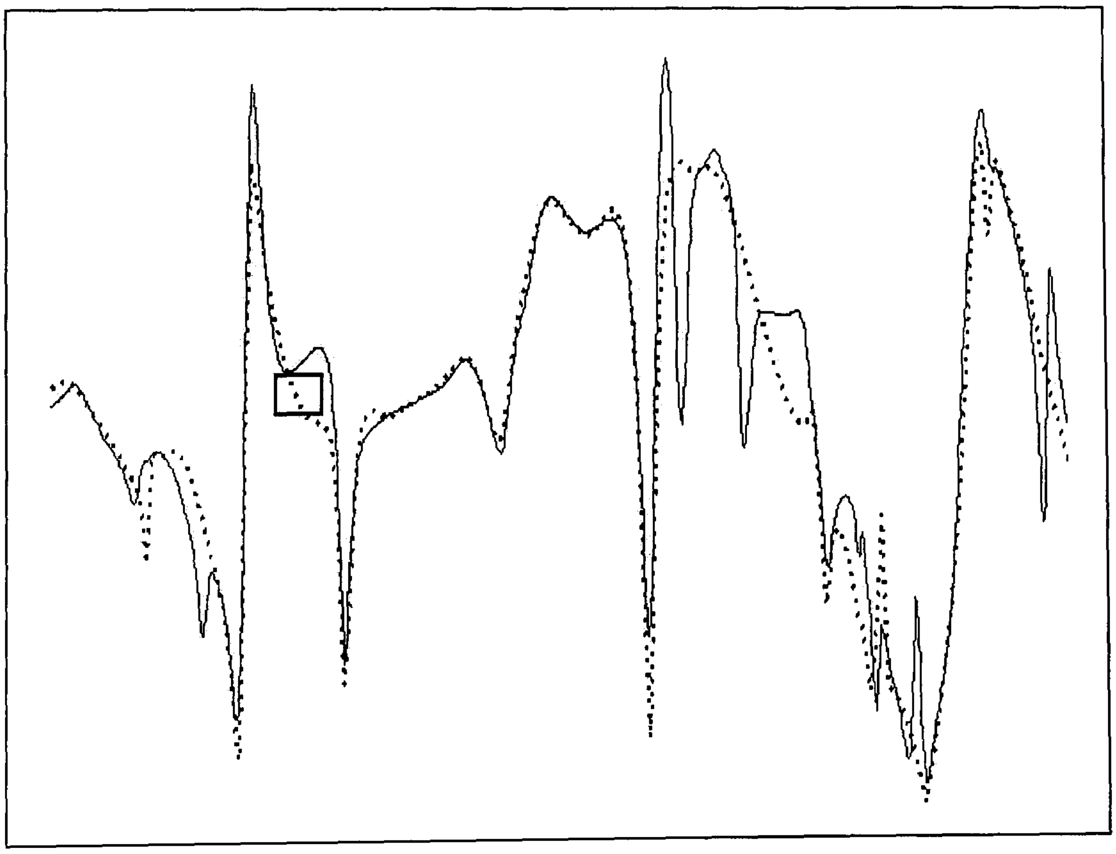

Figure 4.1: Complex data signal comparison

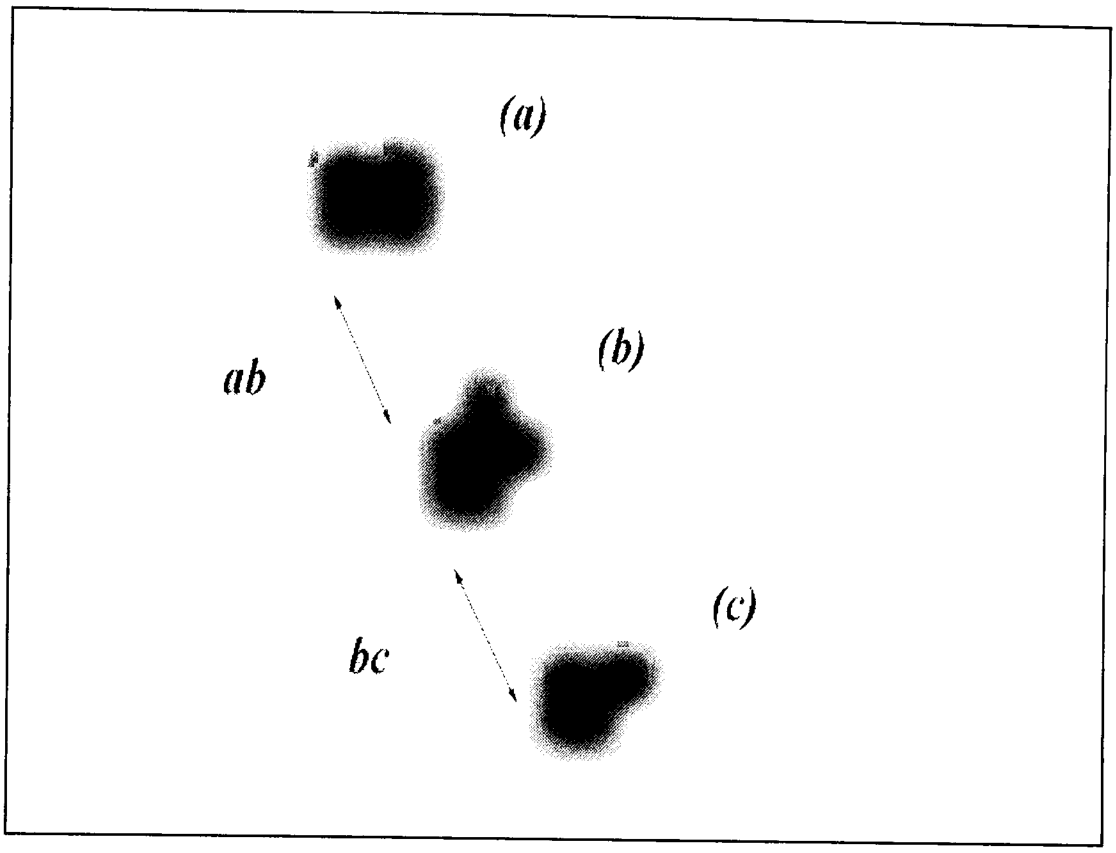

Figure 4.2: Atomic, relational and positional characteristics

However in order to attain a clear representation of atomic, relational and positional characteristics inherent in complex data signals, the algorithms employed to extract critical features must analyse homogenous regions of the signals which directly correspond to the specific extraction algorithm applied. The extraction of gradients and features within any automated validation scheme must be based on data extracted from the full data set containing, exclusively, trend and feature information respectively. It is to this end that Fast Fourier Transforms (FFTs) are employed to remove all information from a signal, except that being used by the dedicated extraction algorithm in question. Furthermore this must be an automatic process, which removes unwanted attributes from the signals before allocating the correct emphasis algorithm. Hence, low frequency component information, or low pass filtered data is employed during the extraction of amplitude and broad feature (trend) information, whilst high frequency component information or high pass filtered data is employed during the extraction of narrow peaks/troughs or features.

Only after isolated and emphasised signatures expressing the atomic, relational and positional attributes of two complex signals are realised, can the process of evaluating discrepancies begin. The inherent characteristics obtained during the signature analysis of two data sets may be viewed as approximated representations of the mental models employed during a visual evaluation of compared signal sets. Employing these numerical signatures, and a specific set of difference algorithms, assessments are made on the quality of compared signals. Through the employment of difference algorithms (and not similarity measures such as those used in correlation), diagnostic representations of the full spectra of validation results are obtained. This step allows for the application of higher order analyses such as statistics to be applied to the validation results of two or more complex signal sets.

The development of a quantitative validation method which produces significant amounts of information expressing the quality of compared data sets, must include a clearly defined qualitative interpretation scale. This scale employs the common categories used by engineers and scientists alike, allowing direct interpretations of quantitative values extracted from the validation procedure. In this way, validation results are presented in a clear manner which relate to the language employed by subjects performing visual evaluations.

4.2 DEVELOPMENT OF THE FEATURE SELECTIVE VALIDATION METHOD

In assessing the quality of a comparison between two data signals, the Feature Selective Validation (FSV) method[Williams 1998,1999] employs two complimentary difference algorithms:

Amplitude discrepancy ***- Amplitude Difference Measure (ADM)***

Feature discrepancy ***- Feature Difference Measure (FDM)***

The rationale for these algorithms is based on a simplified model of the visual search mechanisms employed by engineers undertaking visual comparisons of data signals (detailed in Section 2.2). These may be approximated by a series of absolute, first, and second order derivative differences emphasising atomic, relational and positional information. During a comparison of two data signals, both the ADM and FDM are employed to gain single measurements which indicate differences in amplitude levels/positions (atomic/positional) or differences between feature shapes/positions (relational/positional) respectively. These two complementary measurements are combined by means of vector addition to form a ***Global Difference Measure (GDM)*** which provides a single value representing the overall (atomic, relational and positional) difference between compared data sets. The GDM takes into account differences in the overall amplitude levels of the signals and discrepancies in the location, height or depth, and shapes of trends and features.

Amplitude/Feature discrepancy - Global Difference Measure (GDM)

A further enhancement to the FSV method is the inclusion of optional weighting factors, allowing engineers and scientists to subjectively weight the component measures (ADM and FDM) combined to form the GDM. The weighting factors, namely: the ***Amplitude Difference Tolerance (ADT)*** and the ***Feature Difference Tolerance (FDT)*** allow users to directly weight the component measures of the FSV method based on subjective judgement related to the inherent properties of the compared signals and/or the diverse characteristics of the application area in question. Furthermore, the value of both the ADT and FDT must be greater than zero, and for unbiased assessments, both tolerances are set to unity. It should be noted that, under normal operation, the values of both the ADT and FDT should be chosen so that their multiplication equates to unity (ADT x FDT = unity), allowing a balanced scaling methodology. However, diverse applications may require unbalanced scaling values (detailed in Section 4.3.3), and it is to this end that an automatic balancing algorithm for the weighting tolerances is not implemented in the validation scheme.

In order to produce accurate measures which assess differences between data signals, distinct processing stages are required. Essentially, these analysis stages comprise individual filters and algorithms each performing unique tasks on the raw data and the subsequent data produced by the preceding stages of the scheme. The FSV method uses two distinct processing stages in developing a true representation of differences between compared signals.

The first processing stage within the FSV method employs fourier analysis to filter and isolate homogeneous regions of the full signals in question. The goal of signal segmentation is to separate signals into homogeneous regions so that subsequent algorithms may operate on critical sections of data in an isolated manner, minimising or removing any danger of spurious masking effects. This allows the removal of unwanted areas of the full signals under investigation, which may not relate to the particular processing stage employed at that instant (discussed in detailed in Section 5.4.4).

4.2.1 Signal Segmentation - Unmasking Critical Features

Within a comparison of complex signals, two main regions of the full signals - $I_{SET1}(f)$ and $I_{SET2}(f)$ - must be isolated and analysed independently. These two homogenous regions are the low frequency - $I_{LOW1}(f)$, $I_{LOW2}(f)$ - and high frequency - $I_{HIGH1}(f)$, $I_{HIGH2}(f)$ - components of the full signals respectively. Such that:

$$I_{LOW1}(f) \Leftarrow F_{SET1}(g_I) \qquad g_{I\,\min} + g_{I\,dc} \le g_I < g_{I_L} \qquad \textbf{(4.1)}$$

$$I_{LOW2}(f) \Leftarrow F_{SET2}(g_{II}) \qquad g_{II\,\min} + g_{II\,dc} \le g_{II} < g_{II_L} \qquad \textbf{(4.2)}$$

$$I_{HIGH1}(f) \Leftarrow F_{SET1}(g_I) \qquad g_{I_L} \le g_I \le g_{I\,\max} \qquad \textbf{(4.3)}$$

$$I_{HIGH2}(f) \Leftarrow F_{SET2}(g_{II}) \qquad g_{II_L} \le g_{II} \le g_{II\,\max} \qquad \textbf{(4.4)}$$

where

$$f_{\min} \leq f \leq f_{\max}$$

$$g_{I_L} = \frac{1}{2.5} \sum_{g_{I\min}}^{g_{I\max}} F_{SET1}(g_I)$$

$$g_{II_L} = \frac{1}{2.5} \sum_{g_{II\min}}^{g_{II\max}} F_{SET2}(g_{II})$$

where g_L, g_{dc} and g_{max} denote the low pass cut-off, d.c. and maximum frequency components in the fourier domain respectively, and f_{min} and f_{max} denote the first and last samples in the data signals respectively. Furthermore, $\Leftarrow$ denotes inverse fourier transformation, and $F_{SET1}(g)$ and $F_{SET2}(g)$ denote the fourier domain representations of $I_{SET1}(f)$ and $I_{SET2}(f)$ respectively. It should also be noted that linear ramp filters were employed at both minima and maxima cut-off points in the fourier domain before inverse transformation was applied, with this step removing unnecessary artefacts from the transformed signals.

Figure 4.3 illustrates the comparison of two complex data signals, namely; $I_{SET1}(f)$ and $I_{SET2}(f)$, whilst Figures 4.4 and 4.5 illustrate comparisons of the low frequency $I_{LOW1}(f)$ and $I_{LOW2}(f)$ and high frequency $I_{HIGH1}(f)$ and $I_{HIGH2}(f)$ components of $I_{SET1}(f)$ and $I_{SET2}(f)$ respectively.

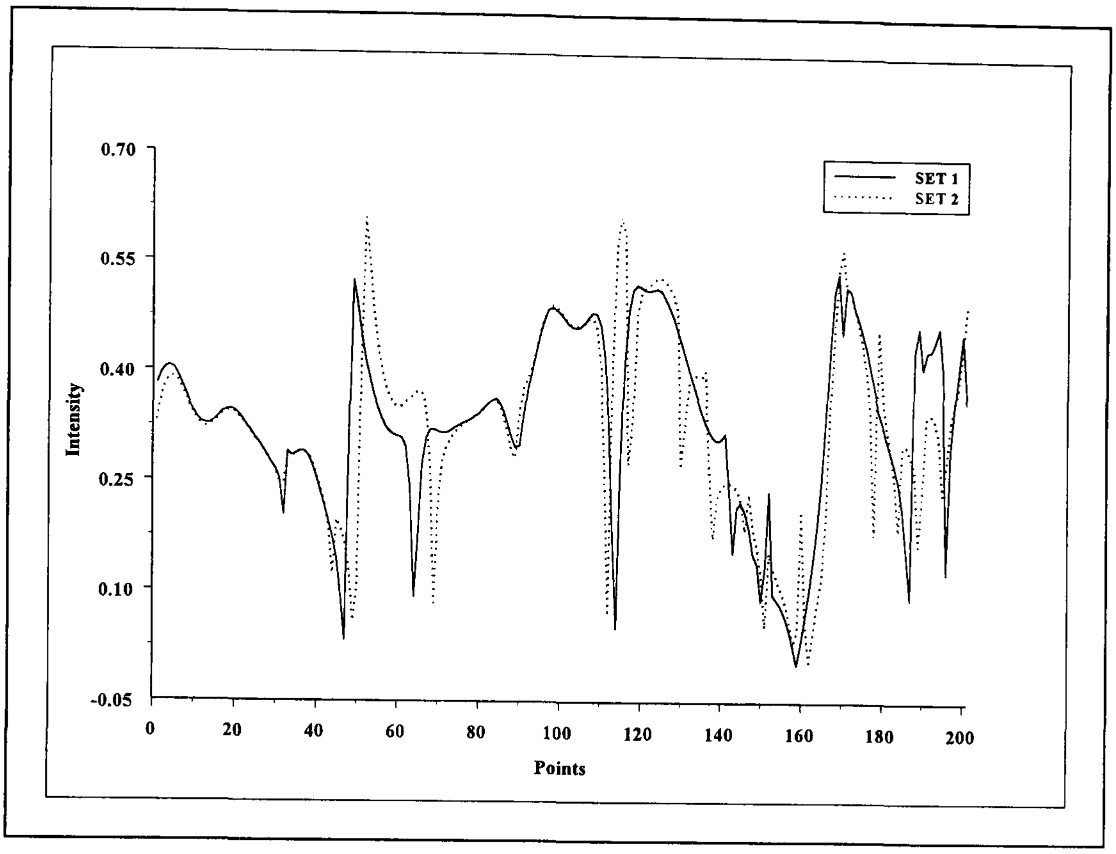

Figure 4.3: Data Sets - $I_{SET1}(f)/I_{SET2}(f)$

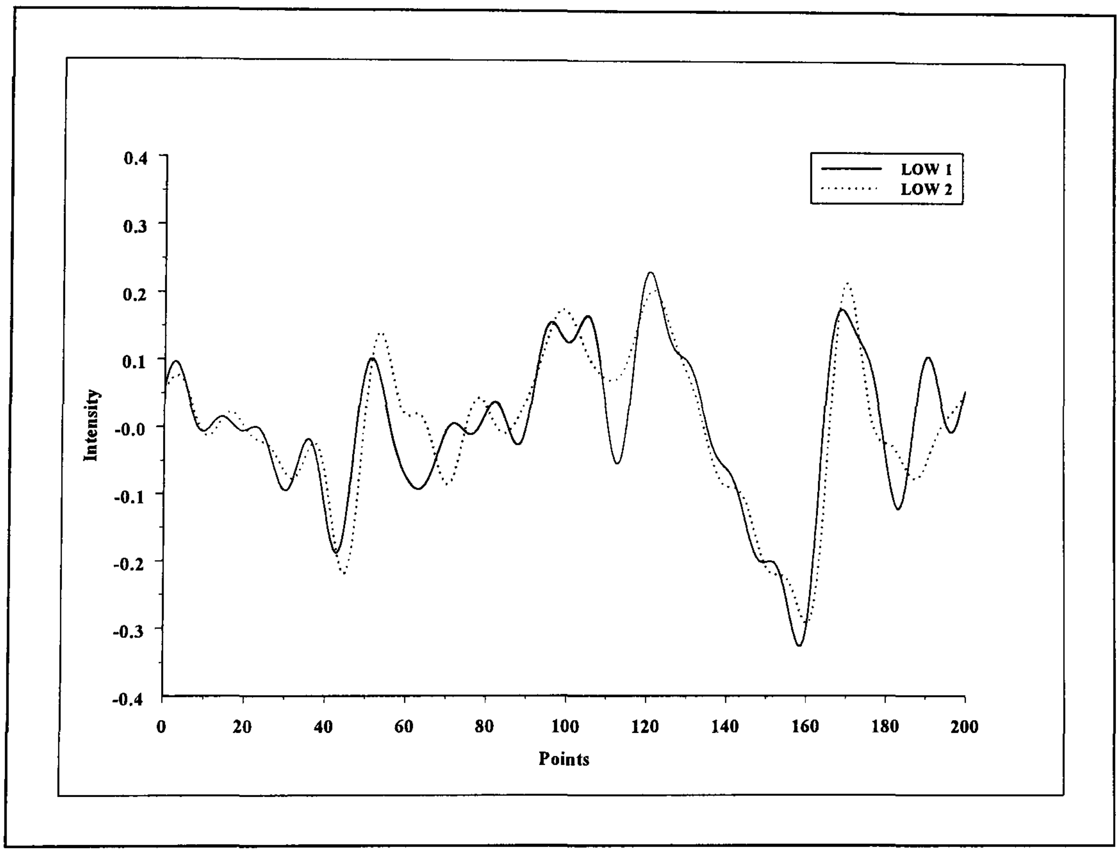

Figure 4.4: Data Sets - $I_{LOW1}(f)/I_{LOW2}(f)$

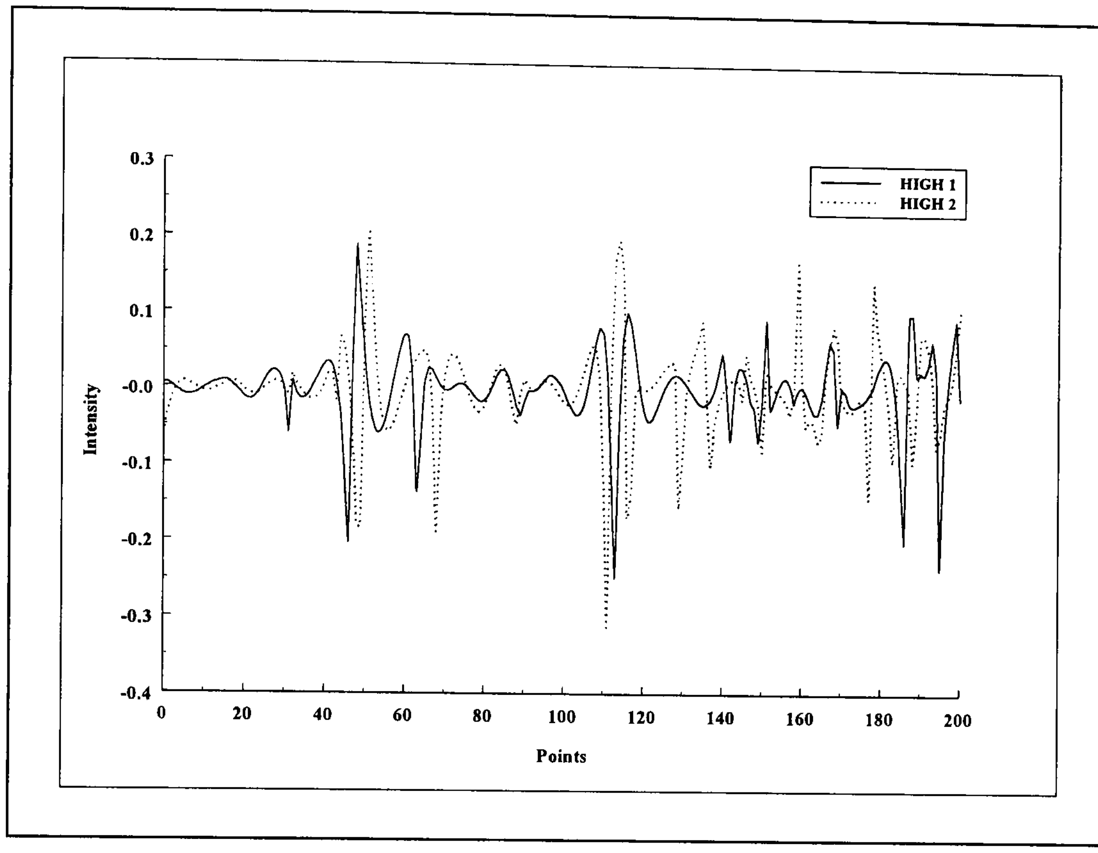

Figure 4.5: Data Sets - $I_{HIGH1}(f)/I_{HIGH2}(f)$

4.2.2 Figures of Merit - An Initial Assessment

The second stage within the FSV method employs the filtered data signals $I_{LOW1}(f)$, $I_{LOW2}(f)$, $I_{HIGH1}(f)$ and $I_{HIGH2}(f)$ derived in Section 4.2.1 and the emphasis/extraction analysis discussed in Section 4.1, along with difference algorithms to produce the two component measures embedded in the FSV method, namely; the ADM and FDM.

4.2.2.1 Amplitude Difference Measure

The Amplitude Difference Measure (ADM) given in Equation 4.5 invokes a single level difference measure based on the absolute values of the low pass filtered data derived in Section 4.2.1. Essentially the ADM is the normalised difference between the low pass responses $I_{LOW1}(f)$ and $I_{LOW2}(f)$. The low pass filter is applied to remove high Q features (narrow peaks/troughs) from the compared signals, allowing a clear evaluation of the relative differences in general amplitude levels/positions between compared signals without the presence and possible masking effects of the high frequency component or features. Fundamentally the ADM mirrors an analysis of atomic differences between compared signals employed in a visual comparison of data signals. This measure does not take into account relational differences between neighbouring atomic features, indicating an evaluation of amplitude level/position differences exclusively.

$$ADM = \sum_{f\,min}^{f\,max} ADT.AD_I(f) \tag{4.5}$$

where

$$AD_I(f) = \frac{|I_{LOW1}(f) - I_{LOW2}(f)|}{\alpha_{AD_I}} \tag{4.6}$$

$$\alpha_{AD_I} = \frac{1.5}{f\,max - f\,min} \sum_{f\,min}^{f\,max} |I_{LOW1}(f)| + |I_{LOW2}(f)| \tag{4.7}$$

where $AD_I(f)$ denotes the Amplitude Difference sub-measure response illustrated in Figure 4.6 and α_{AD_I} denotes the amplitude normalisation factor, equated to the average absolute energy contained in the signals under investigation. The *ADT* scales the *ADM* value included in the *GDM*, and is given a value of unity for unbiased operation. The inclusion of the scaling factor 1.5 in the evaluation of α_{AD_I} scales the amplitude difference measure to the required quantitative values indicated in Table 4.1 described in Section 4.2.3. This scaling methodology was conceived employing a substantial amount of combined visual evaluation feedback from highly trained experts in several fields of study.

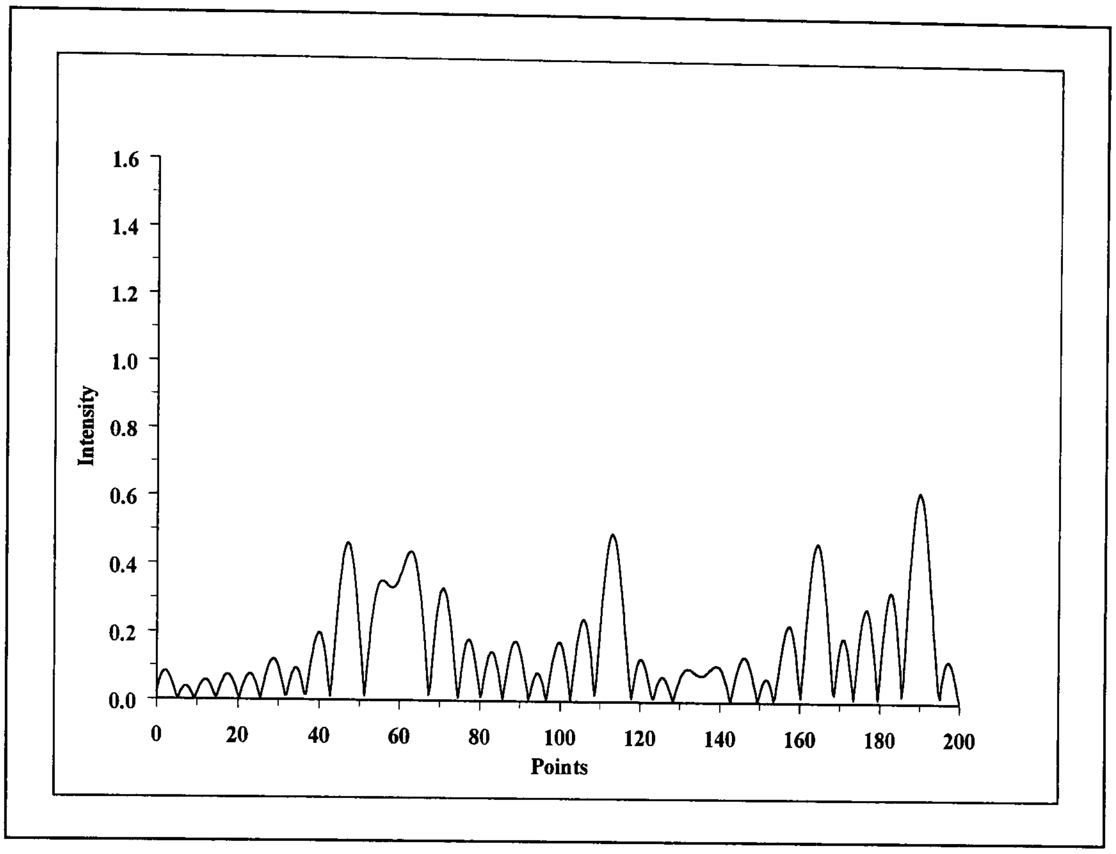

Figure 4.6: ***$AD_I(f)$*** **response curve - Equation 4.6**

The importance of this measure increases when scrutinising signals possessing very high resonant behaviour i.e. peaks and troughs (it is important to note that the low and high frequency components within a signal are not necessarily related - e.g. signals containing high levels of noise). In the case where a signal contains high levels of noise, it is imperative to gain a measure of the true - reduced noise - differences between compared signals. In order to gain a noise limited comparison, both the *ADT* and *FDT* are reduced to values less than unity, emphasising the assessment of amplitude and trend differences, and de-emphasising differences between features.

4.2.2.2 Feature Difference Measure

The Feature Difference Measure (FDM) defined in Equation 4.8 invokes three sub-level difference algorithms, namely: Feature Difference I ($FD_I(f)$), Feature Difference II ($FD_{II}(f)$) and Feature Difference III ($FD_{III}(f)$). Each of these sub-level difference measures emphasise independent areas of the compared signals, producing complementary feature difference information.

The $FD_I(f)$ sub-measure given in Equation 4.9 comprises the normalised first derivative difference between the low pass responses $I_{LOW1}(f)$ *and* $I_{LOW2}(f)$ derived previously. Essentially the $FD_I(f)$ sub-measure illustrated in Figure 4.7 emphasises low order trend differences between compared signals, allowing an accurate low level analysis of the shapes and positions of broad features. Emphasis is placed on the differences in instantaneous gradients between the compared signals, allowing an assessment of differences between feature shapes to be realised. As this measure is equated employing indexed samples, the value of this measure increases as shifts between broad features increase, indicating positional differences between features. This measure may be viewed as a low level gradient detection, emphasis and difference algorithm. The $FD_I(f)$ sub-measure mirrors relational/positional differences between neighbouring atomic features (broad features) employed during a visual comparison of signals.

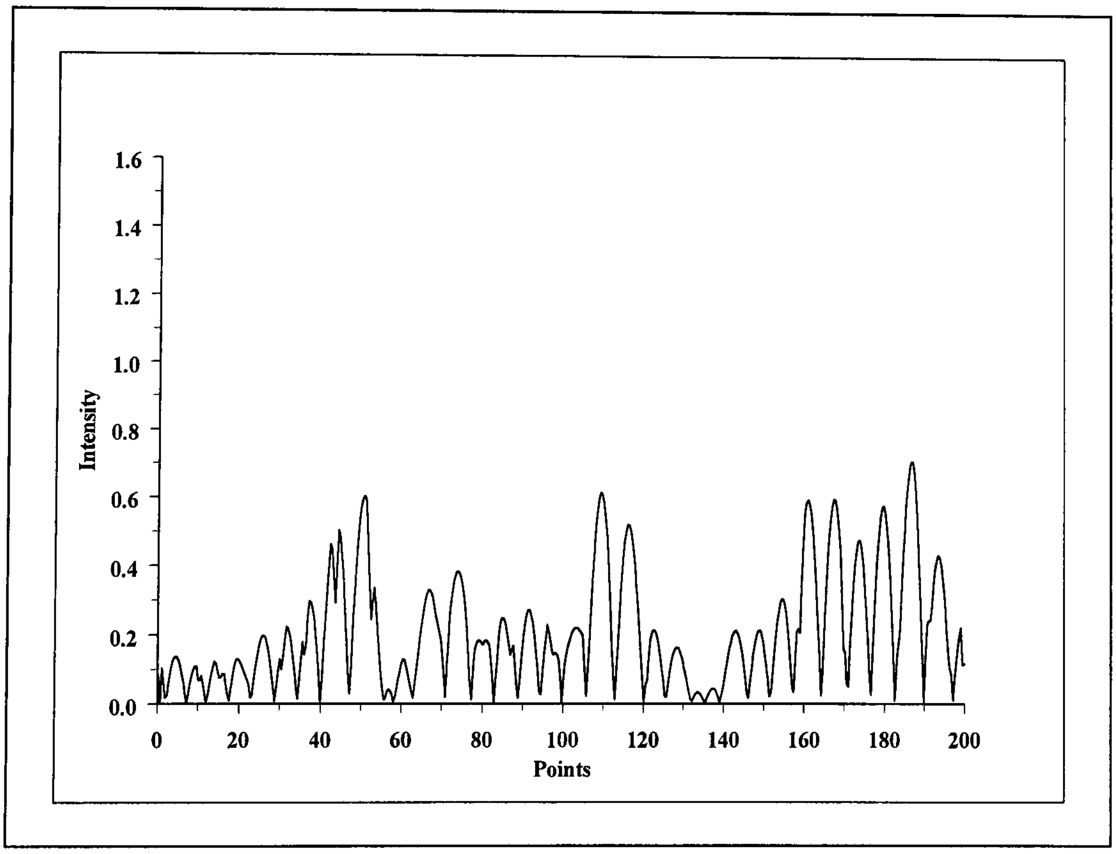

Figure 4.7: $FD_I(f)$ response curve - Equation 4.9

The $FD_{II}(f)$ sub-measure given in Equation 4.11, similar in structure to that of Equation 4.9 employs the normalised first derivative difference between the high pass responses $I_{HIGH1}(f)$ and $I_{HIGH2}(f)$ derived previously. The $FD_{II}(f)$ sub-measure illustrated in Figure 4.8 emphasises higher order trends, allowing an accurate low level analysis of differences between the shapes and positions of narrow features. The $FD_{II}(f)$ sub-measure mirrors low level relational/positional differences between neighbouring atomic features (narrow features) employed during a visual comparison of signals.

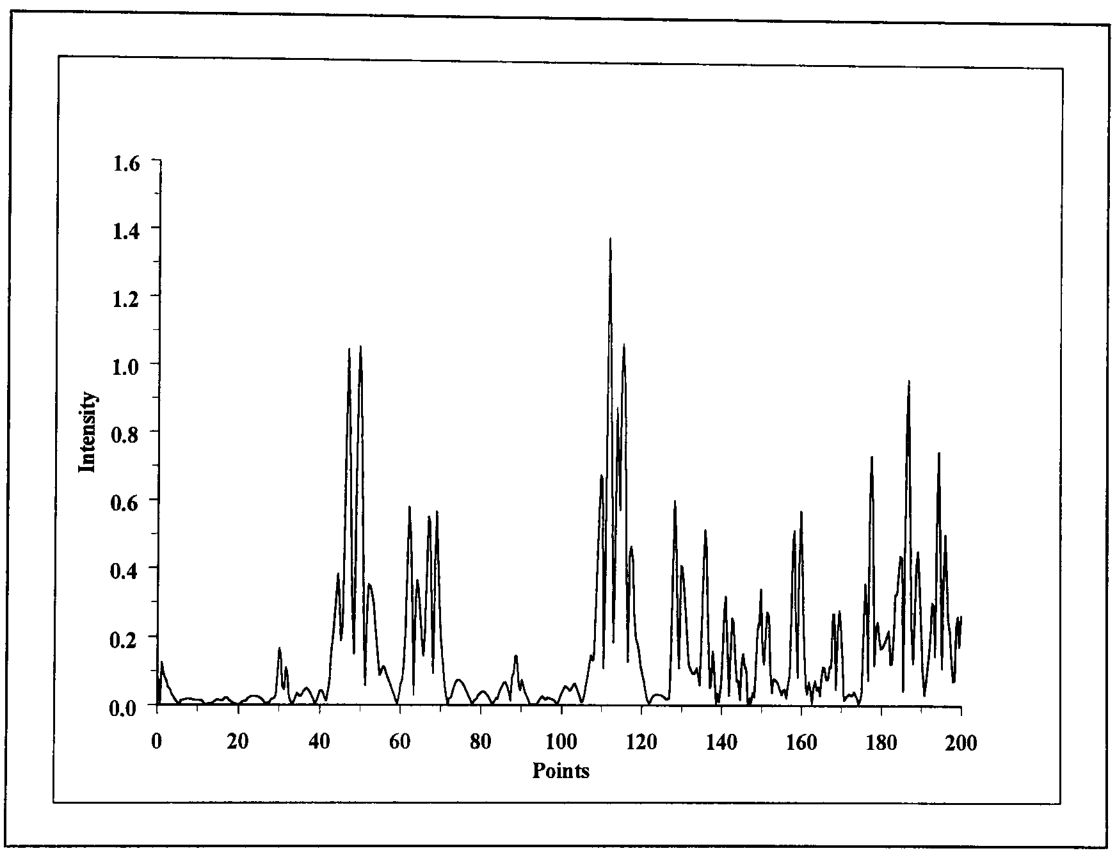

Figure 4.8: *$FD_{II}(f)$* response curve - Equation 4.11

The final sub-measure $FD_{III}(f)$ given in Equation 4.13, comprises the normalised second derivative difference between the high pass responses $I_{HIGH1}(f)$ and $I_{HIGH2}(f)$. The $FD_{III}(f)$ sub-measure illustrated in Figure 4.9 emphasises higher order features, allowing an accurate high level analysis of differences between the shapes and positions of narrow features. This measure is fundamentally similar to the $FD_{II}(f)$ sub-measure, but increases the emphasis placed on differences in the positions of high level features in a comparison. This measure may be viewed as an edge detection, emphasis and difference algorithm. The $FD_{III}(f)$ sub-measure mirrors high level relational/positional differences between neighbouring atomic features (narrow features) employed during a visual comparison of signals.

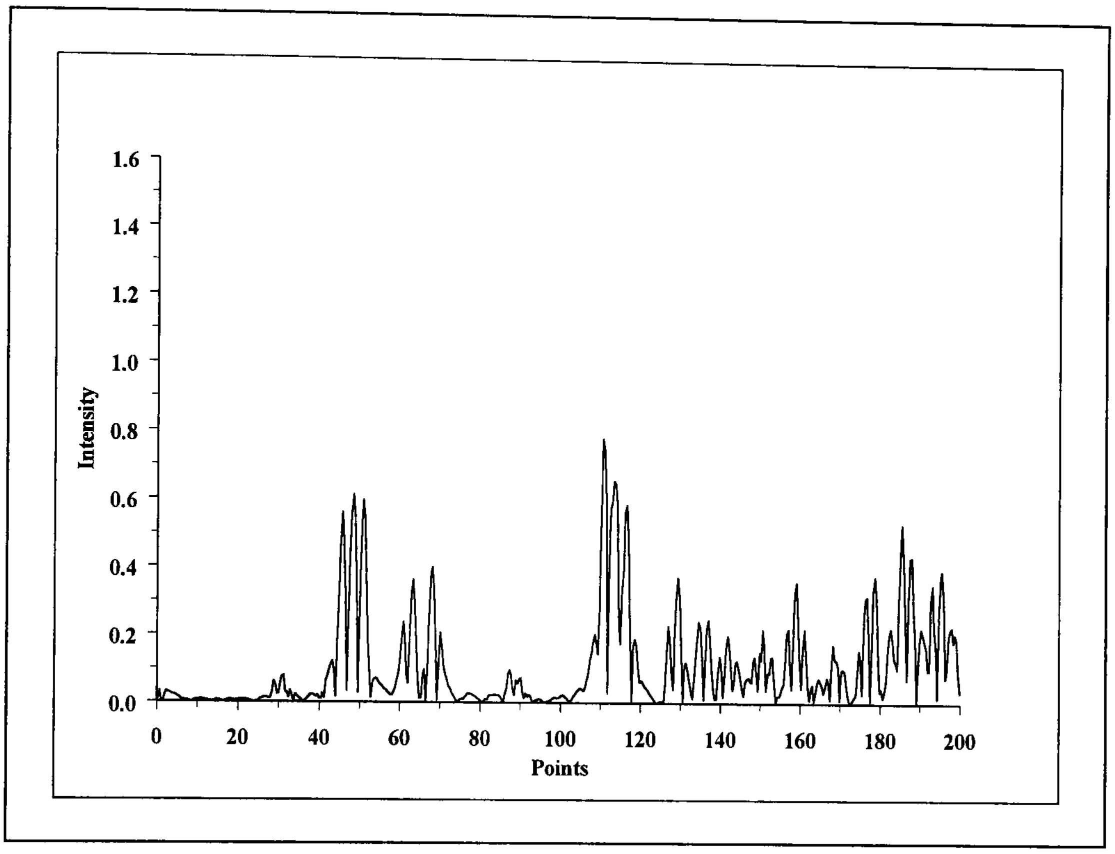

Figure 4.9: $FD_{III}(f)$ response curve - Equation 4.13

All three sub-measures are integrated into the final Feature Difference Measure (FDM) given in equation 4.8, employing the second subjective value presented by the user, namely, the Feature Difference Tolerance (FDT). The FDT value operates on the FDM in a two fold manner; firstly balancing the internal sub-measures, emphasising either low level trends (broad peaks/troughs) or higher level features (narrow peaks/troughs). Secondly the FDT value scales the FDM value included in the evaluation of the GDM. An FDT value of less than unity emphasises differences in low level trends, whilst relegating the importance of high level features (which may be associated with noise).

$$FDM = \sum_{f\,\min}^{f\,\max} \frac{FDT}{\beta_{III}} \cdot \left(\frac{FD_I(f)}{\beta_I} + FD_{II}(f) + \frac{FD_{III}(f)}{\beta_{II}} \right) \qquad \textbf{(4.8)}$$

where

$$FD_I(f) = \frac{|I'_{LOW1}(f) - I'_{LOW2}(f)|}{\alpha_{FD_I}} \qquad \textbf{(4.9)}$$

$$\alpha_{FD_I} = \frac{2}{f\max - f\min} \sum_{f\min}^{f\max} |I'_{LOW1}(f)| + |I'_{LOW2}(f)| \qquad \textbf{(4.10)}$$

$$FD_{II}(f) = \frac{|I'_{HIGH1}(f) - I'_{HIGH2}(f)|}{\alpha_{FD_{II}}} \qquad \textbf{(4.11)}$$

$$\alpha_{FD_{II}} = \frac{4}{f\max - f\min} \sum_{f\min}^{f\max} |I'_{HIGH1}(f)| + |I'_{HIGH2}(f)| \qquad \textbf{(4.12)}$$

$$FD_{III}(f) = \frac{|I''_{HIGH1}(f) - I''_{HIGH2}(f)|}{\alpha_{FD_{III}}} \qquad \textbf{(4.13)}$$

$$\alpha_{FD_{III}} = \frac{6}{f\max - f\min} \sum_{f\min}^{f\max} |I''_{HIGH1}(f)| + |I''_{HIGH2}(f)| \qquad \textbf{(4.14)}$$

where f_{min} and f_{max} denote the lowest and highest components in the data set respectively, $I_{SET1}(f)$ and $I_{SET2}(f)$ denote the comparison data signals respectively, and all equations are normalised to the sum of the average absolute energy contained in their respective signals forming intelligent normalisation factors. The single and double primes represent first and second derivatives respectively and FDT denotes the Feature Difference Tolerance described previously. The scaling factors (2, 4 and 6) included in equations 4.10, 4.12 and 4.14 respectively allow the quantitative values obtained from the feature difference measure to be scaled employing Table 4.1, with the magnitude of these scaling factors derived employing the combined visual evaluation results of highly trained subjects from several fields of study. Furthermore, β_I and β_{II} denote internal balancing mechanisms employed to weight the component measures within the FDM.

For an FDT value of unity

$$\beta_I = \beta_{II} = Unity \tag{4.15}$$

For all FDT values less than or greater than unity

$$\beta_I = \sqrt{FDT} \quad \beta_{II} = \frac{1}{\sqrt{FDT}} \tag{4.16}$$

Furthermore, β_{III} denotes a re-balancing value which normalises the balance values applied to the sub-measures within the FDM algorithm of equation 4.8.

$$\beta_{III} = \frac{12}{2\beta_I + 4 + 6\beta_{II}} = \frac{Unbiased \quad values}{Biased \quad values} \tag{4.17}$$

The FDM is essentially an analysis of maxima, minima, gradients and inflexions within the compared signals. This measure accounts for both feature positions and feature shapes, whilst relegating the importance of the absolute difference in amplitudes between features. This allows for an assessment of the importance of individual features based on their width, magnitude, position and overall shape.

4.2.2.3 Global Difference Measure

Both the ADM and the FDM have been developed to produce independent information based on the measures employed in visual evaluations of results undertaken by engineers or scientists alike. However, it is important in any automated validation scheme to produce from these sub level measurements, a measure expressing the overall quality of a comparison. This is accomplished through vector addition of the ADM and FDM which form the Global Difference Measure (GDM) given in equation 4.18. The GDM gives a clear indication of both amplitude and feature differences between compared signals, quantifying the overall assessment of a comparison. Essentially, an unbiased evaluation of the GDM mirrors the atomic, relational and positional differences taken into consideration during a visual comparison of data signals.

$$GDM = \sum_{f_{\min}}^{f_{\max}} \sqrt{\left(ADT.AD_I(f)\right)^2 + \left(\frac{FDT}{\beta_{III}}\left(\frac{FD_I(f)}{\beta_I} + FD_{II} + \frac{FD_{III}(f)}{\beta_{III}}\right)\right)^2} \quad \textbf{(4.18)}$$

4.2.3 Informative Scaling - Associated Quality Bands

A principle difficulty in attempting the development of meaningful automated validation schemes, is that of producing accurate yet informative statistical information about the quality of a comparison. Furthermore, the information gained from these statistical analyses must be presented in a categorical manner which mirrors 'human' interpretations. The GDM and its constituent measures (ADM and FDM) embedded in the FSV method produce single quantitative figures of merit which express the quality of a comparison. In order to be understood, these quantitative values must be qualitatively defined employing a conventional scale employed by engineers and scientists alike. Within the FSV method, the qualitative assessment of the GDM and its component measures indicates that a quantitative value of 0 represents an ideal comparison, whilst increasing quantitative values represent increasingly poor comparisons. It should be noted that the category boundaries lie midway between the central values indicated in the expanded interpretation scale of Table 4.1:

FSV Intensity (Quantitative)	*Interpretation (Qualitative)*
0.00	***Ideal***
0.05	***Excellent***
0.10	***Very good***
0.20	***Good***
0.40	***Fair***
0.80	***Poor***
1.60	***Extremely poor***

Table 4.1: FSV interpretation scale

It should be noted, that the GDM value obtained from an unweighted FSV analysis will never be less than the smallest value of the two component measures (i.e. ADM and FDM), due to the nature of the GDM algorithm. The ADM, FDM and GDM figures of merit for the comparison illustrated in Figure 4.3 are indicated in Table 4.2. The GDM is equated to 0.42 or 'fair', the FDM indicates significant discrepancies between data signals and is equated to 0.37 or 'fair', whilst the ADM is equated to 0.12 indicating that the compared amplitude levels are in 'very good' agreement. It is clear from the example indicated in Table 4.2 that discrepancies between feature shapes and positions (FDM) constitute a major part of the overall (GDM) error between the data signals illustrated in Figure 4.3. It is conjectured - for illustration - at this point that, acceptable quality may be set at a GDM value not greater than 0.2 (i.e. 'good'). The choice of this value is dependent on the inherent sensitivity of the method of data acquisition, and the application area under investigation. Within the FSV method, this value of GDM acceptability is defined as the ***Global Difference Tolerance*** (***GDT***).

Measure	***Quantitative value***	***Qualitative value***
ADM	***0.12***	***Very good***
FDM	***0.37***	***Fair***
GDM	***0.42***	***Fair***

Table 4.2: Figures of merit - $I_{SET1}(f)/I_{SET2}(f)$

However, the mean values (figures of merit) produced by each of the individual measures (ADM, FDM and GDM) of the FSV method provide little information concerning the frequency content of a validation response curve. Whilst the figures of merit do give a true indication of the disagreement between comparison signals, they do not uniquely categorise a particular validation result, as, in general, a number of validation response curves may share the same mean value. This in no way undermines the need for such mean evaluations within the scheme, however it does indicate that further information must be provided to strengthen the meaning and uniqueness of these measures if required.

4.2.4 Confidence Levels - Categorising Comparisons

Consolidation of this inadequacy in the global figures of merit is provided in the form of confidence levels for each of the figures of merit produced by the FSV method. Whilst the evaluation of these confidence levels is a relatively simple task, the information contained in the measures is highly valuable, providing, in part, a level of uniqueness to each comparison. The adoption of these confidence levels has illustrated a powerful tool in accurately analysing and categorising comparisons of complex signals (detailed in Chapter 5). In essence each confidence level comprises the percentage of a comparison response curve which falls into each of the qualitative validation criteria's indicated in Table 4.1, mirroring the combined category effects of engineers and scientists undertaking visual evaluations discussed in Section 2.5.

Within this scheme the confidence level labelled 'good' for the ADM constitutes the percentage of points along the ADM response curve (***ADM(f)***) which fall into the 'good' criteria (0.15 to 0.3), whilst the confidence level associated with the category of 'extremely poor' for the GDM constitutes the percentage of points along the GDM response curve (***GDM(f)***) which fall into the 'extremely poor' criteria (1.2 or above). The confidence levels included in the FSV method may be viewed as a statistical extension to the analysis applied to a comparison.

Figures 4.10, 4.11, and 4.12 illustrate the ADM, FDM and GDM confidence levels respectively, obtained from the comparison illustrated in Figure 4.3. The ADM confidence levels indicate 'very good' agreement between amplitude levels, however considerable confidence is contained in the neighbouring quality bands. This indicates that whilst the majority (28%) of amplitude levels are in 'very good' agreement, 16% of the amplitude levels are in 'ideal' agreement, and 17% of amplitude levels only attain 'fair' agreement. The FDM confidence levels indicate 'fair' agreement between features, however the confidence levels are widely dispersed, covering all seven quality bands. This indicates that whilst the majority (27%) of features are in 'fair' agreement, only 1% of the features are in 'ideal' agreement, but 6% of features attain 'extremely poor' agreement. Despite the widely dispersed GDM confidence levels, the comparison illustrated in Figure 4.3 can be confidently placed in the 'fair' quality band indicating an average comparison.

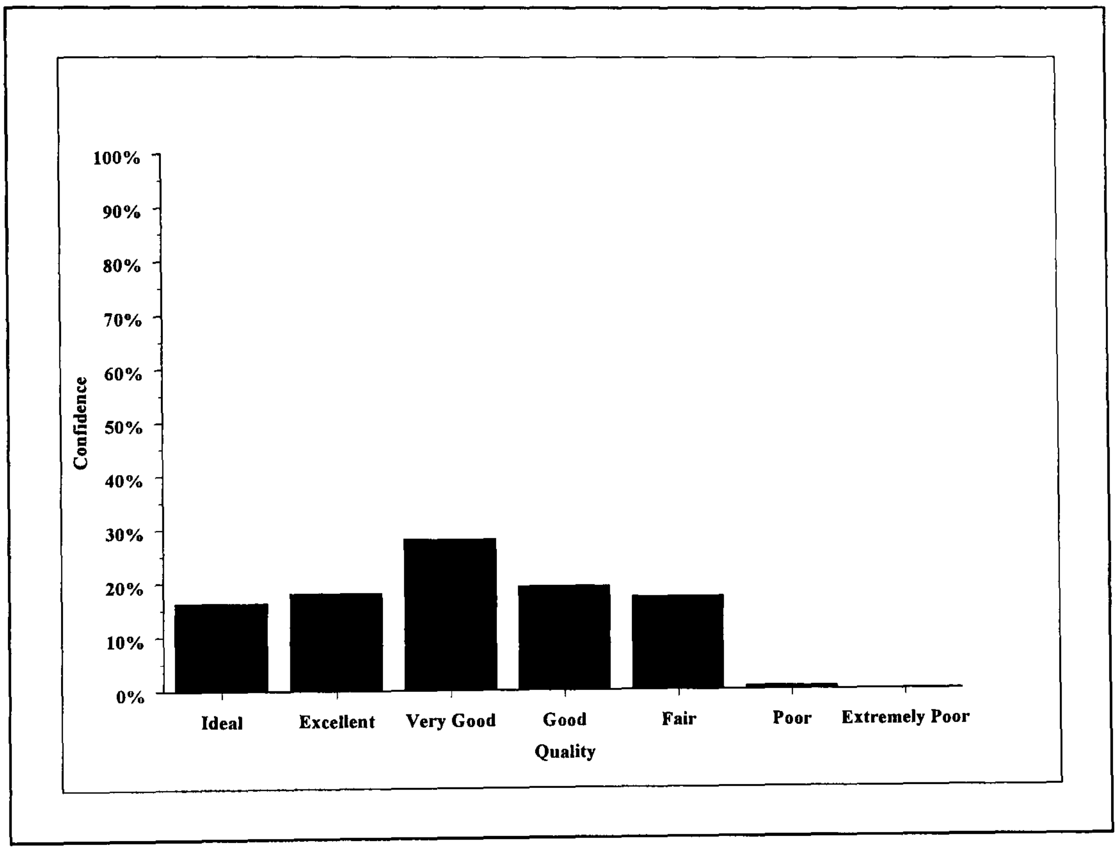

Figure 4.10: ADM confidence levels - $I_{SET1}(f)/I_{SET2}(f)$

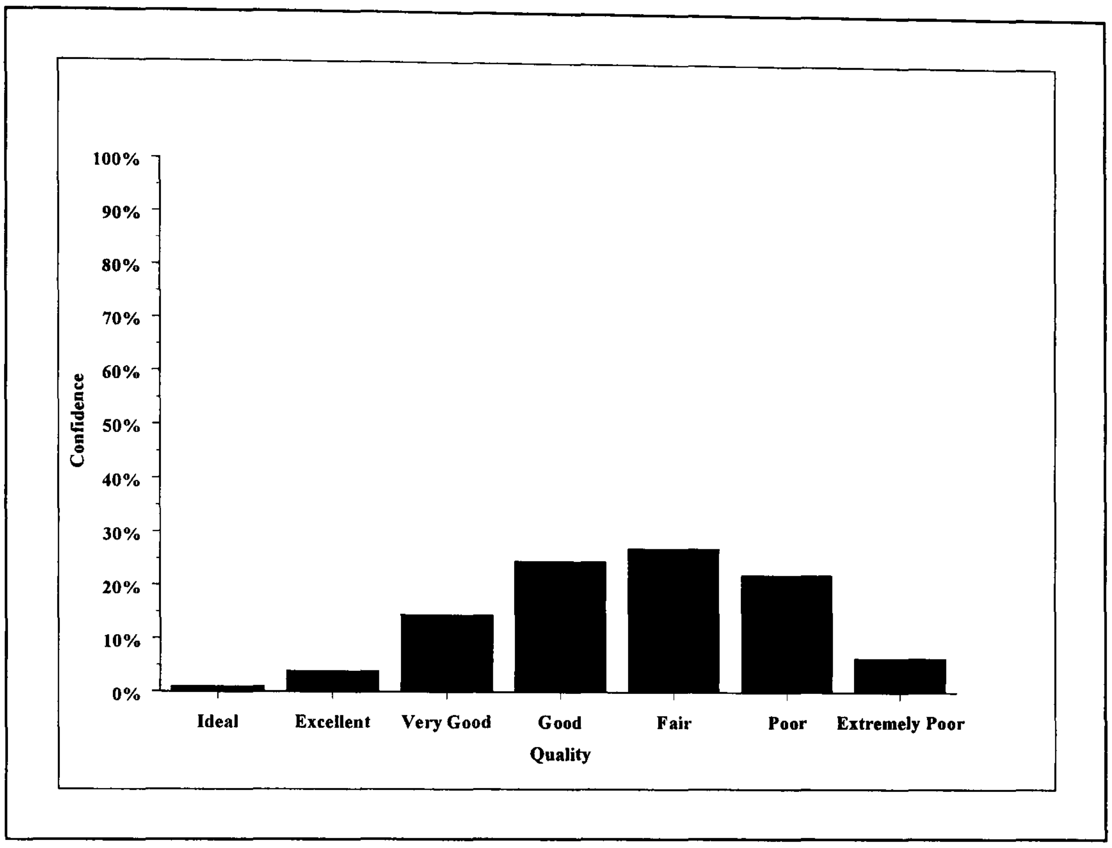

Figure 4.11: FDM confidence levels - $I_{SET1}(f)/I_{SET2}(f)$

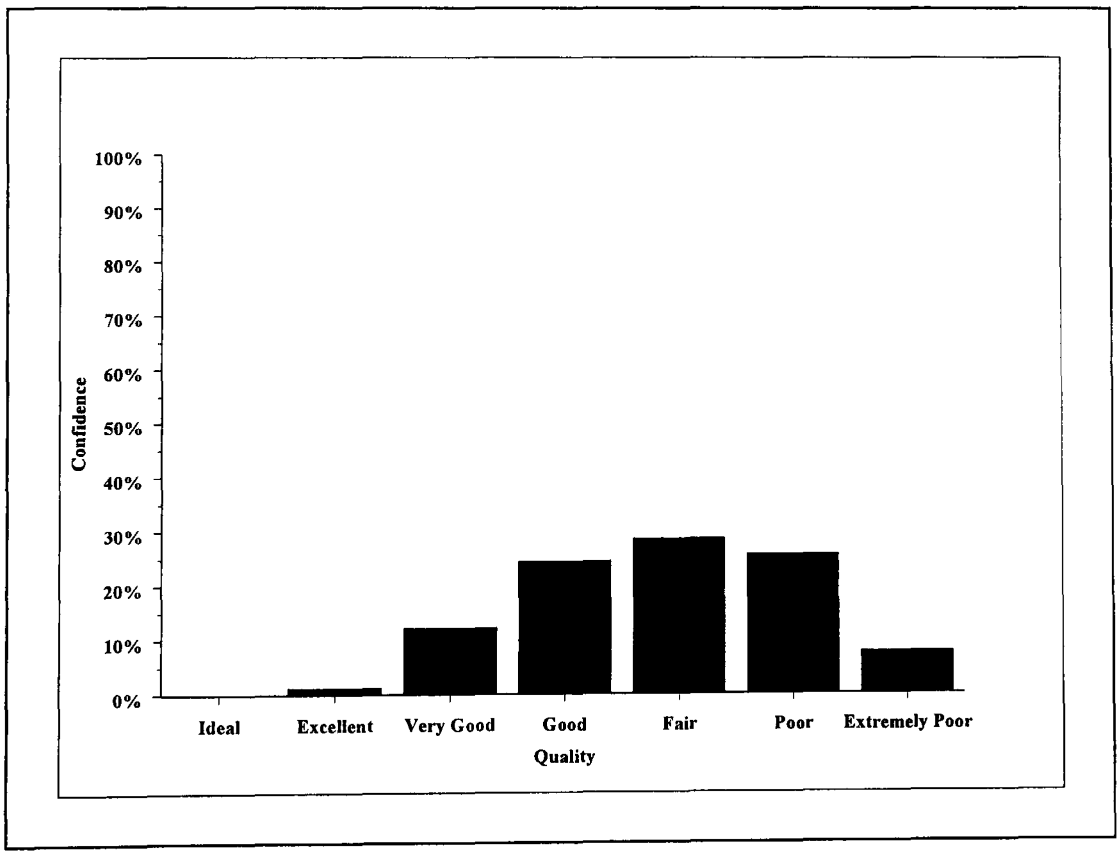

Figure 4.12: GDM confidence levels - $I_{SET1}(f)/I_{SET2}(f)$

4.2.5 Diagnostics - In-depth Analysis

Employment of both the figures of merit and confidence levels included in the FSV method allow a clear indicator of the nature of a comparison, and serve as the first level of information within the FSV scheme. Furthermore, the ADM, FDM and GDM produce single assessment figures based on the quality of a comparison. However, by omitting the summations from Equations 4.5, 4.8 and 4.18, an assessment figure for each point on the compared signals can be achieved for each of the individual measures. This allows the realisation of confidence levels and graphical representations of each of the three measures embedded in the FSV method, allowing a highly detailed diagnosis of the comparison in question.

Figures 4.13, 4.14 and 4.15 illustrate the graphical representations of the *ADM(f)*, *FDM(f)* and *GDM(f)* discrete response curves respectively. Figure 4.13 indicates amplitude level discrepancies centred around six main points within the compared data signals, namely: 45; 60; 115; 165; and 190. Whilst Figures 4.14 and 4.15 illustrate significant discrepancies between both the shape and positions of resonant features within three main regions of the compared data signals, namely: 40 - 60; 100 - 120; and 160 - 200.

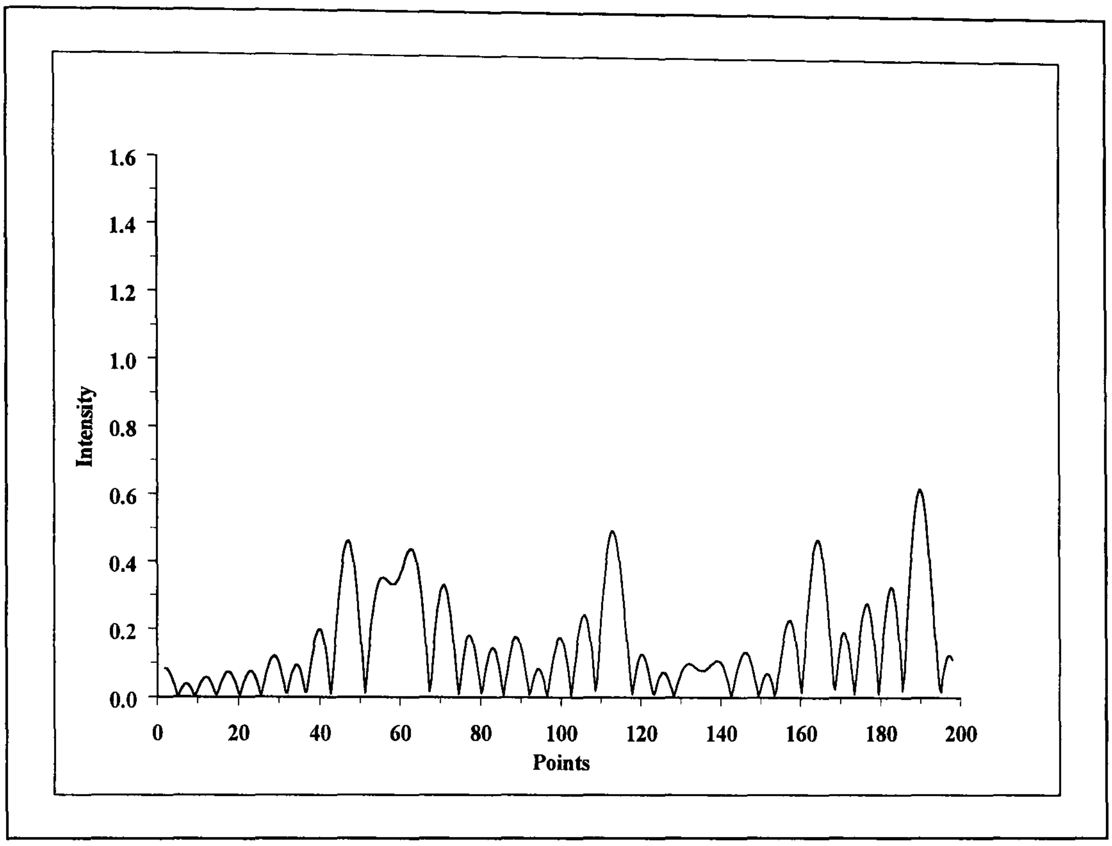

Figure 4.13: Amplitude Difference Response - *ADM(f)*

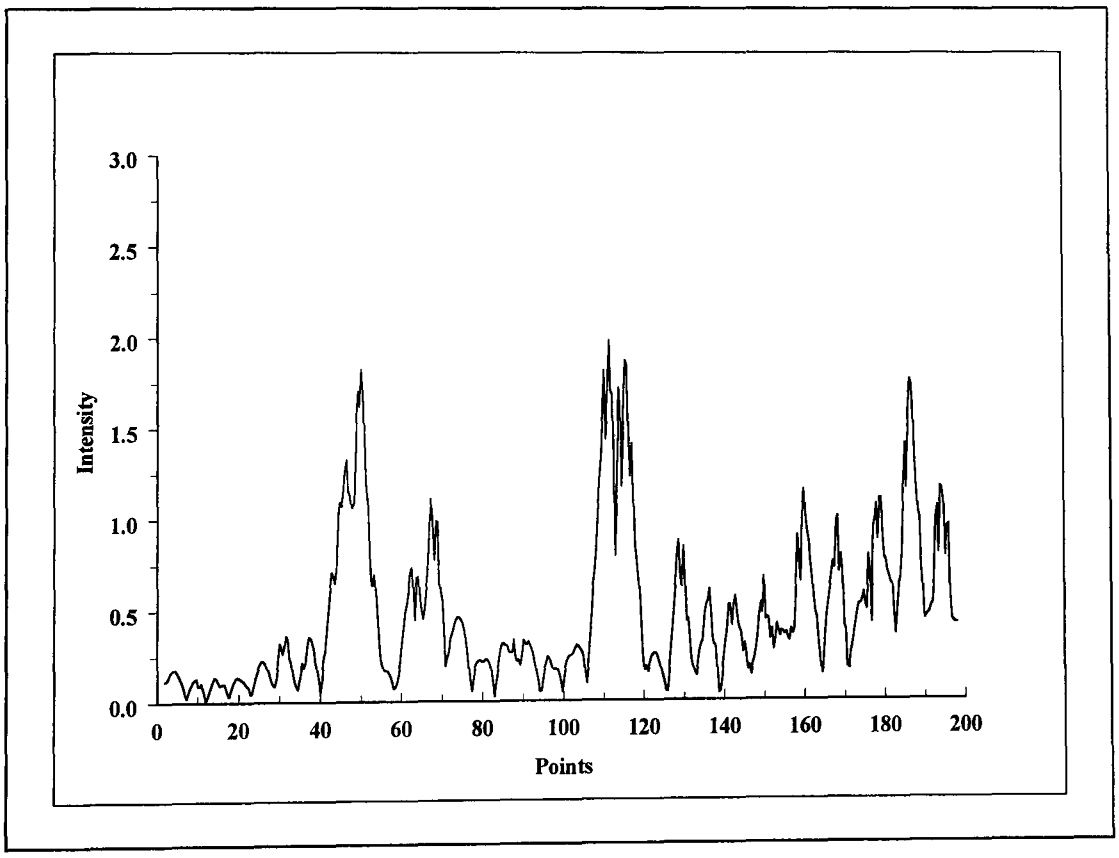

Figure 4.14: Feature Difference Response - *FDM(f)*

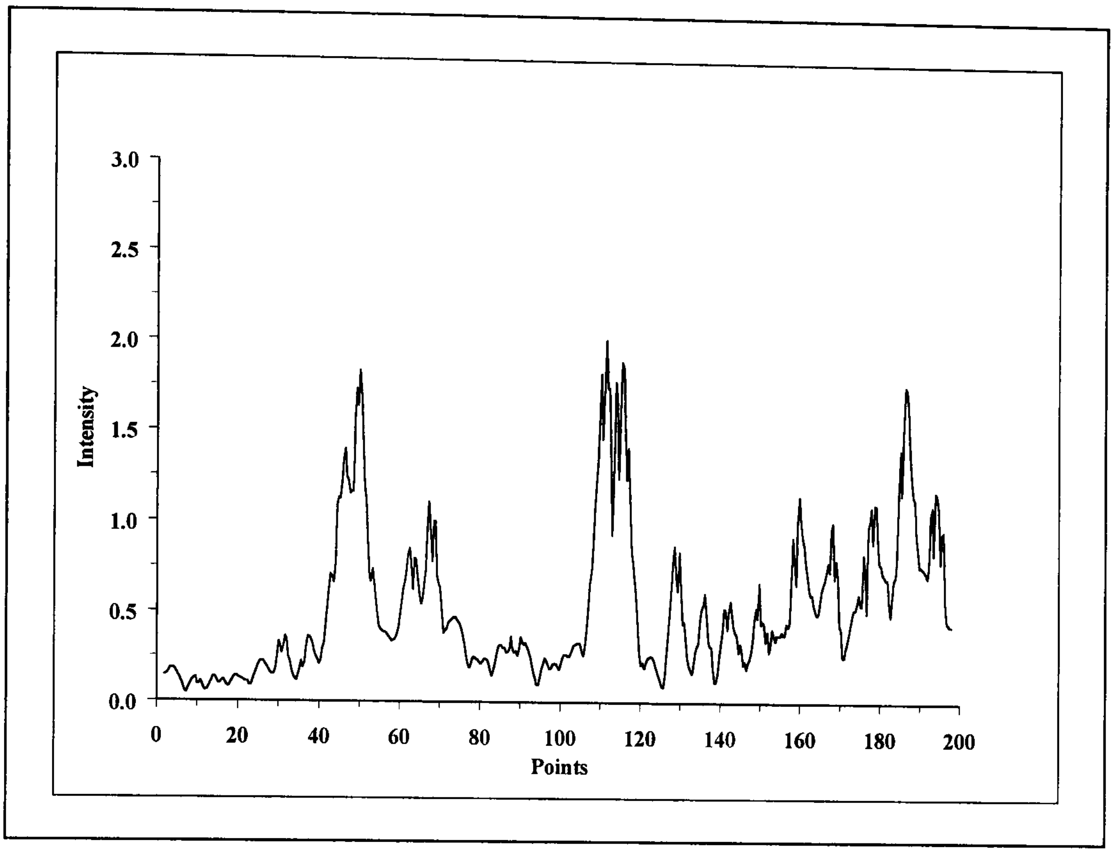

Figure 4.15: Global Difference Response - *GDM(f)*

The graphical representations of each component measure within the FSV method allow engineers and scientists to identify the locations and types of features that provide major contributions to a 'poor' comparison, allowing directed remedial action to be taken if appropriate. This provides a rigorous framework for confidence building and improved experimental or numerical techniques. Furthermore, as the assessment is undertaken quantitatively: a natural progression of rank ordered features which contribute to a 'poor' comparison are presented, encouraging a rigorous and focused investigation into the major sources of error between data sets.

4.2.6 Discussion

The single figures of merit, associated confidence levels and graphical representations extracted from the FSV method can be employed to produce simple pass/fail guidelines, to identify a level of acceptable error, or enable an assessment to be made as to whether differences are due to the method of acquiring results. However, in many application areas comparison data sets will incur distortion effects during data acquisition, or as a direct result of different techniques employed to produce data signal sets. Within the field of Deoxyribonucleic Acid (DNA) analysis for example, samples from the same subject may be distorted if different gel types are employed to extract DNA sequences. In order to accurately assess the discrepancies between DNA sequences, it is first necessary to correct any incurred distortion effects inherent in the comparison data sets. An extension to the FSV method is the development of the Feature Selective Correction (FSC) method. Such distortion correction is an essential pre-processing option to any validation method allowing the correction of distorted signals before an analysis of discrepancies between compared signals is employed.

4.3 SIGNAL CORRECTION

In order to assess the true quality of a comparison within many disciplines, complex distortions between compared signals must first be analysed and corrected. The human visual/perceptual system allows subjects whether consciously or sub-consciously to account for such distortions, and whilst visual evaluation possesses the ability to correct data sets, a valid judgement on the quality of a comparison may or may not be made employing this technique. In order to accommodate a highly complex data correction method within an automated validation scheme, several key processing stages are required.

4.3.1 Current Correction Method

One correction method[Menacer 1992], has shown that linear stretches between distorted signals may be corrected by employing correlation (discussed in detail in Section 3.1). Fundamentally, this algorithm employs two single feature signals; $I_1(f)$ *and* $I_2(f)$ illustrated in Figure 4.16.

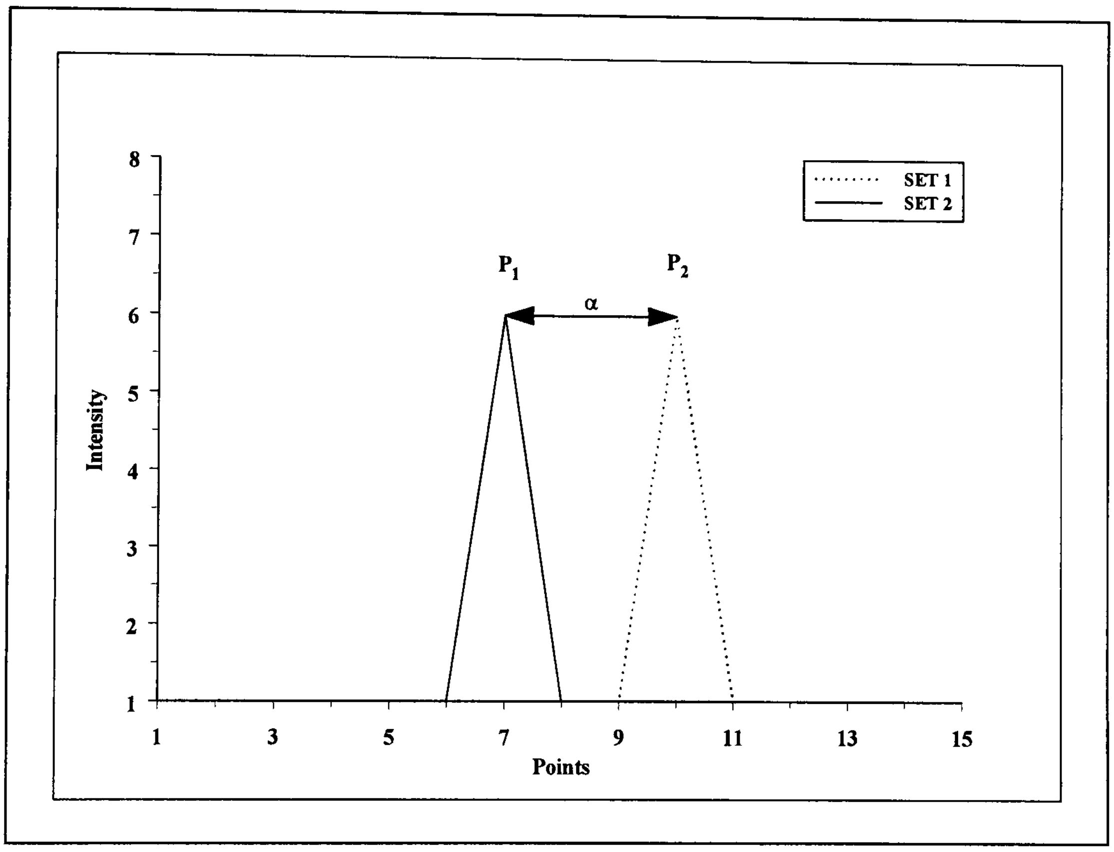

Figure 4.16: Data Sets - $I_1(f)/I_2(f)$

Where $I_1(f)$ and $I_2(f)$ denote target and comparison signals respectively. Furthermore, the distortion between these signals is taken to be linear, such that:

$$P_2 = F.P_1 \tag{4.19}$$

where P_1 and P_2 denote the peak positions embedded in the target and comparison signals respectively. While F is a linear stretching factor defined as:

$$F = \frac{N}{N-\alpha} \tag{4.20}$$

where α denotes an unknown positive integer relating to the stretch between P_1 and P_2. However, the value of F equated in Equation 4.20 underestimates the true value of shift between P_1 and P_2 which is equated to α.

Further equations in the method of Menacer employ the correlation algorithm in the approximation of the integer value α, incorporating this value into the index correction function of Equation 4.21:

$$I_{CORR2}(f) = I_2(\frac{N}{N-\alpha} \cdot f) \qquad \textbf{(4.21)}$$

where α denotes an approximated value of stretch between P_1 and P_2 and $I_{CORR2}(f)$ denotes an under corrected version of the comparison signal $I_2(f)$.

Further iterations of the method must be applied until α is reduced below a predefined value, or the cross correlation coefficient of the compared signals does not increase. However, continuous corrections of distorted signals employing linear stretch mechanisms invariably cause degradation of the original signal. Peaks and troughs embedded in the data signal under correction are partially lost due to the phenomenon of aliasing. In order to minimise this problem, high resolution signals (i.e. over-sampled or super-sampled signals) must be employed during the correction analysis. However, this inevitably increases both the computational overheads and the time taken to perform distorted signal correction.

The method of Menacer operates on a single level, applying stretch mechanisms to the full spectra of the signal $I_2(f)$. However, due to the index multiplication applied in Equation 4.21, data must be discarded from the boundaries of the of the comparison signal, rendering the method unstable in cases where several features are embedded in the compared signals. This phenomenon is illustrated in Examples 4.1 and 4.2.

Example 4.1: Single Feature Correction - Menacer

Figure 4.17 illustrates the comparison of two signals, exhibiting single features embedded in the target and comparison signals respectively. Employing the method of Menacer, an initial offset between target and comparison features α is approximated to 2 points. With the comparison feature leading the target feature.

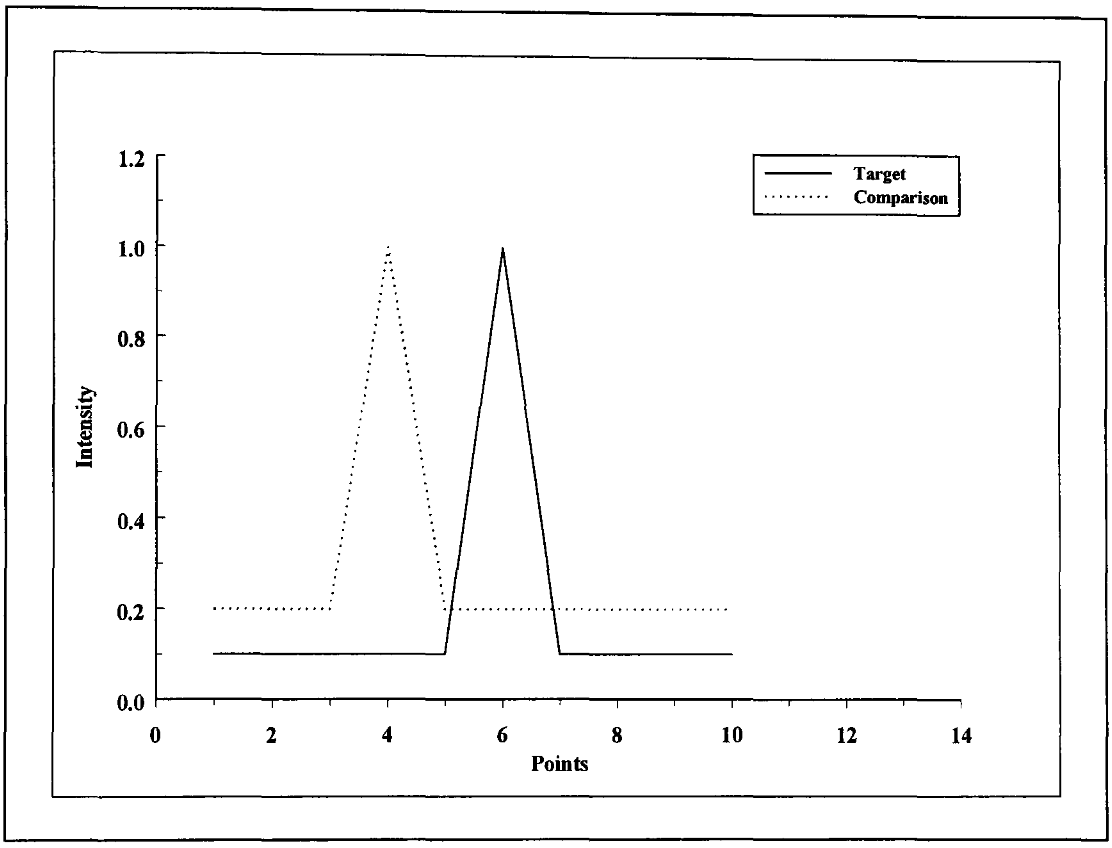

Figure 4.17: Target and comparison signals - single feature

Substituting the value of α into Equation 4.21, where N is equated to 10 (the number of samples in the signal data sets) a partial linear correction of the comparison signal index may be implemented, such that:

$$Corrected(f) = Comparison(\frac{10}{10-2}.f) \qquad f_{\min} \leq f \leq f_{\max}$$

If this analysis is iteratively applied to the comparison signal, a fully corrected version of the comparison signal index may be realised. An approximated version of the corrected comparison signal and the original target signal are illustrated in Figure 4.18, with the offset between target and comparison features removed. However, the multiplication factor applied in Equation 4.21 discards the corrected comparison data which overlaps the original boundaries of the data sets (i.e. samples now indexed greater than f_{max}). Whilst this does not adversely affect the signals within this example, when multiple features are corrected employing this method the resulting signals are effected.

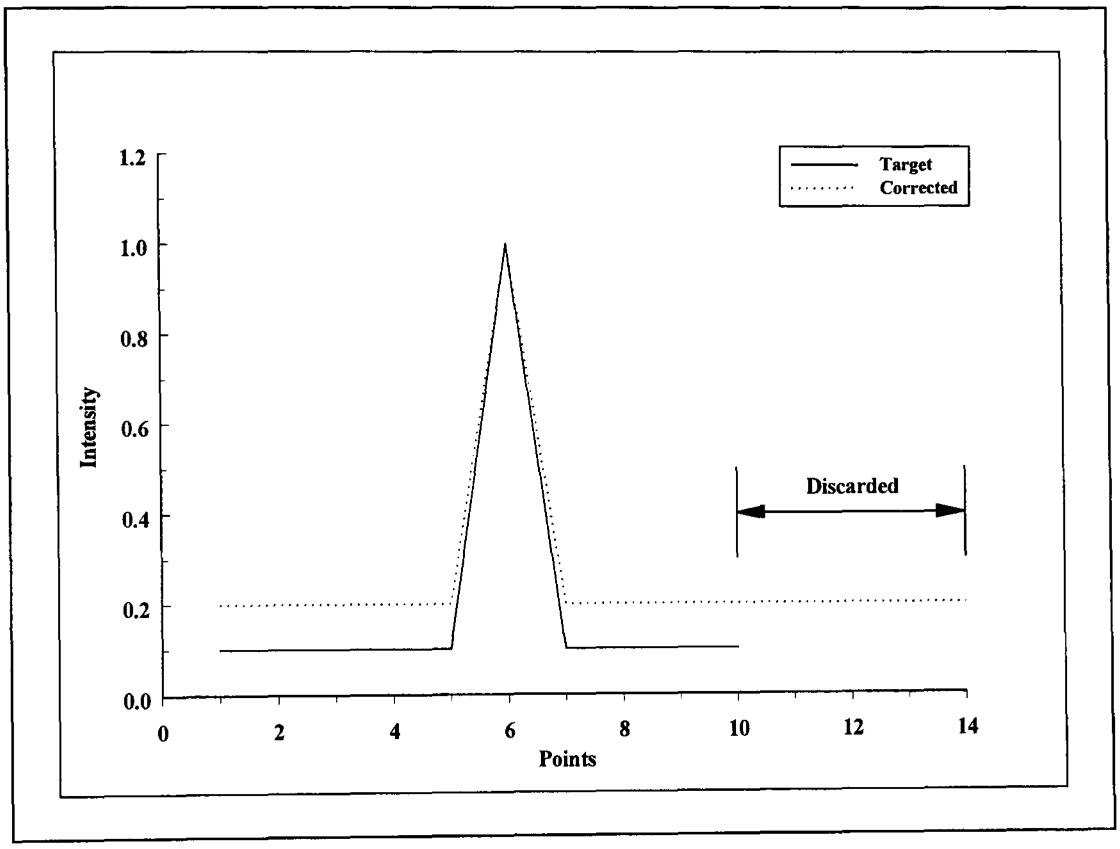

Figure 4.18: Corrected signal - single feature

Example 4.2: Multiple Feature Correction - Menacer

Figure 4.19 illustrates the comparison of two signals, exhibiting multiple features embedded in the target and comparison signals respectively. Employing the method of Menacer, an initial offset between target and comparison features α is approximated to 2 points.

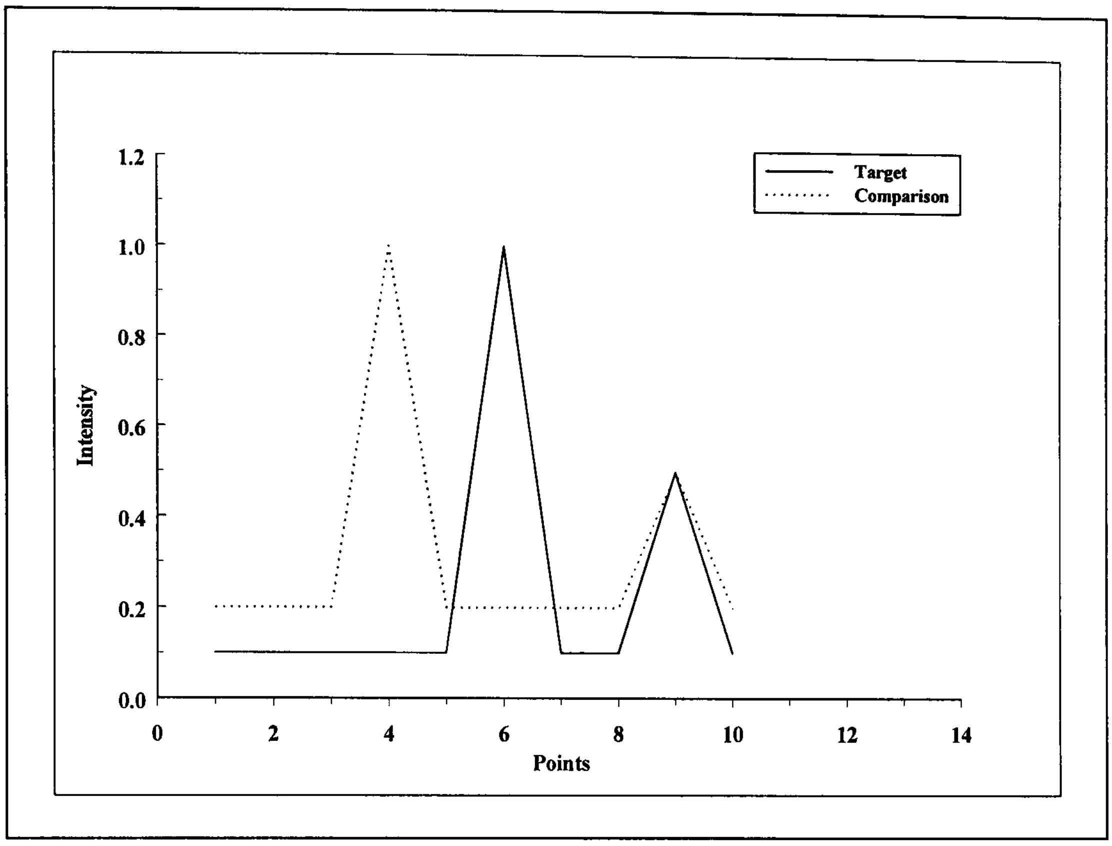

Figure 4.19: Target and comparison signals - multiple features

Figure 4.20 illustrates the corrected multiple feature comparison signal obtained employing the analysis detailed in Example 4.1 and more specifically Equation 4.21. However, employing the method of Menacer for the case detailed in this example where several features are corrected introduces significant errors between the corrected and target signals. Data points discarded from the upper boundary of the original comparison signal include a second feature which (employing a single level correction method) can not be recovered.

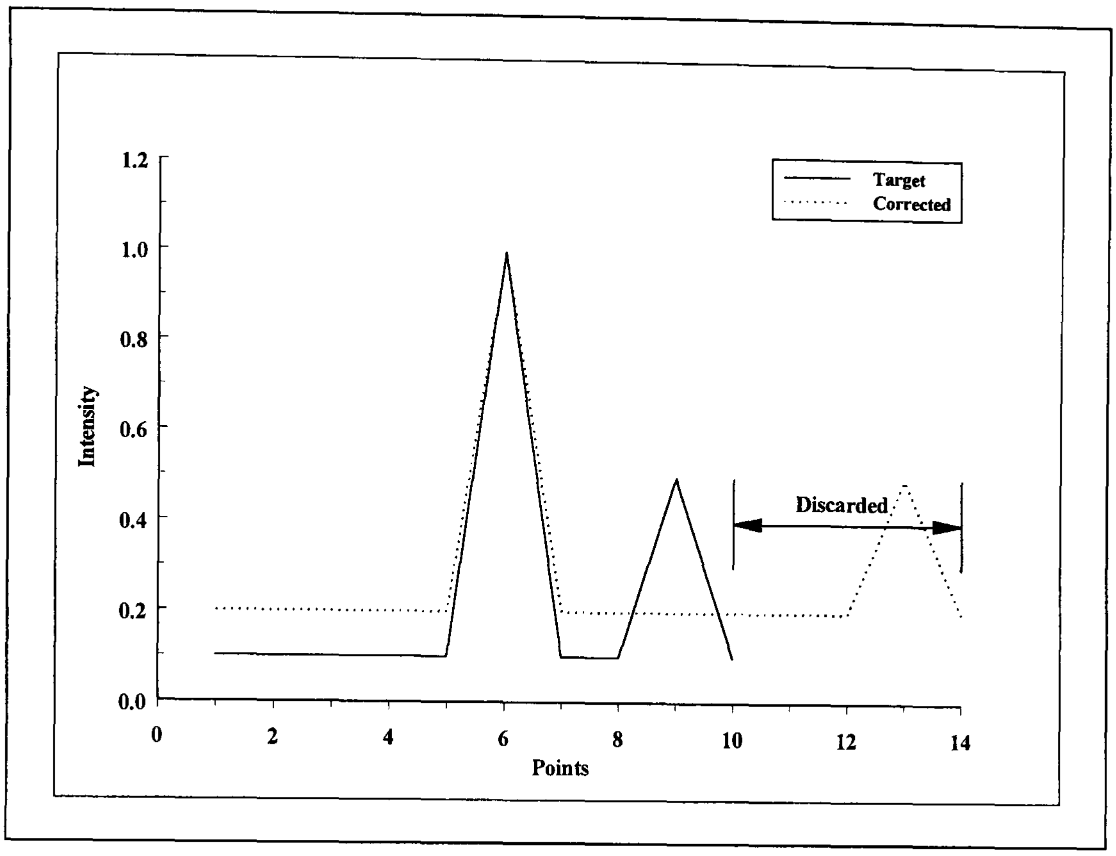

Figure 4.20: Corrected signal - multiple features

4.3.2 Discussion

Whilst the method of Menacer may be employed to correct linear distortions between simple signals, it lacks discernment in cases where highly complex non-linear stretches are embed in the comparison signal. Furthermore, under this scheme, aliasing effects which reduce the size of corrected features are not removed or even minimised. It is to this end that a method of high speed, low loss, complex non-linear correction for distorted signals must be developed. However, this type of analysis is complicated due to the complex nature of the offsets embedded within the signals. These offsets may fall into a mixture of many mathematical categories including: linear, exponential, and logarithmic. The process of taking sections of a comparison signal and iteratively manipulating it applying different mathematical algorithms until it best fits a target signal would be a computationally expensive process requiring huge processing power and speed whilst incurring high costs in both time and equipment.

4.3.3 Development of the Feature Selective Correction Method

It is proposed that linear and non linear distortions embedded within a signal can be corrected employing iterative processing and assessment stages within a decrementing window distortion correction scheme. The processing stages within the proposed method consist of linear stretch mechanisms, and the application of this strategy to reducing sub windows of the full data set allows for the realisation and correction of approximated non-linear stretches (e.g. exponential) within the distorted signal. Furthermore, if an account is held of the distortions corrected throughout a rigorous analysis of distorted signals, single pass interpolation employing this distortion data (advance/delay) can be applied to a high resolution copy of the original signal, greatly reducing the aliasing effects caused by low resolution correction. This overcomes both the problem of the speed at which distorted signals may be corrected, and the accuracy at which the method may be implemented.

The ***Feature Selective Correction*** (***FSC***) method, employs the FSV method, and more specifically the Global Difference Measure (GDM). This value allows a measured level of confidence to be equated which indicates the differences between two signals. Whilst the GDM provides a rigid assessment of the overall differences between compared signals, a measured level of flexibility can be provided through the employment of two optional tolerance values embedded in the FSV method, namely: ADT and FDT. These tolerance values may be viewed as balance mechanisms which bias an assessment towards either the ADM or FDM. A normal or unbiased assessment of a comparison employs unity values for both the ADT and the FDT. However, within a comparison of distorted signals exhibiting high values of noise, the GDM may be biased towards an assessment which emphasises amplitude and trend differences and de-emphasises feature differences, employing tolerance values of less than unity for both the ADM and FDM.

In order to overcome the problems of computational time and cost involved in correcting distorted signals, approximations must be made on the nature of the embedded shifts within the data. The application of linear interpolation routines which stretch the signal, allow linear shifts to be corrected within the full signal being manipulated. A further enhancement to this method employs this analysis applied to windows within the full data set. These windows reduce after each successive stage within the method, allowing the analysis to be applied to smaller and smaller sections of the data. This in turn allows for the realisation of non-linear stretch mechanisms within the manipulation process. The first six stages of the reducing window scheme employed by the ***FSC*** method are illustrated in Figure 4.22, where N denotes the total number of samples contained in the complex signals under investigation, namely: $I_{SET1}(f)$ and $I_{SET2}(f)$ illustrated in Figure 4.21.

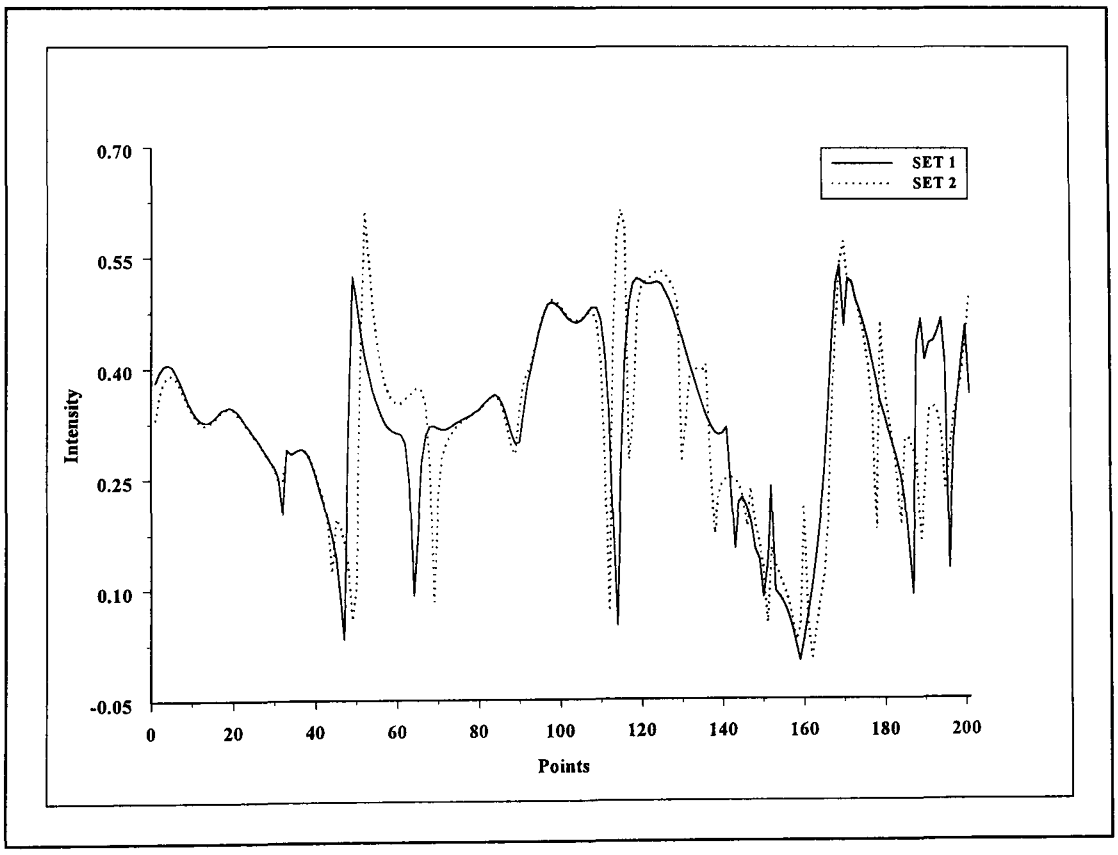

Figure 4.21: Data Sets - $I_{SET1}(f)/I_{SET2}(f)$

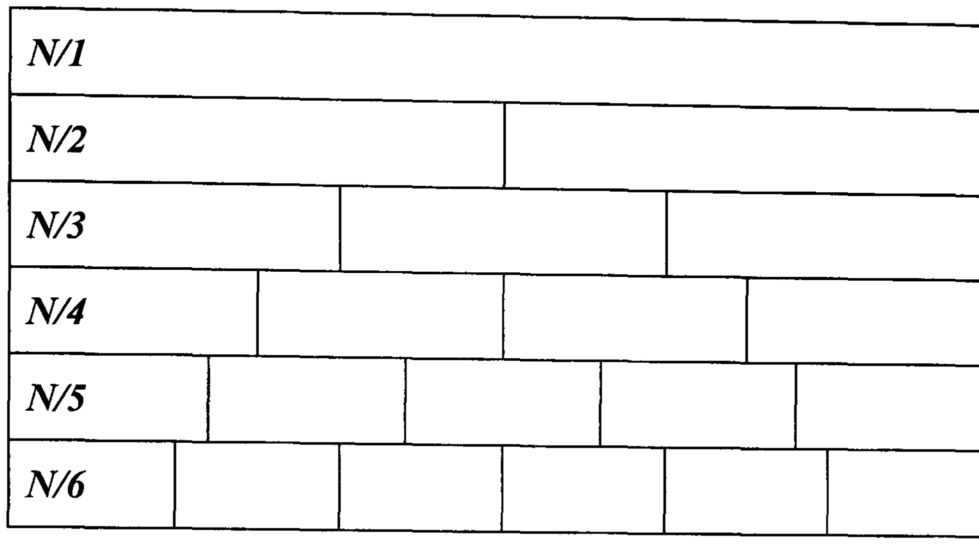

Figure 4.22: Reducing windows scheme employed by the FSC method

There are several automated search schemes which may incorporate the correction mechanism detailed previously. The most common of which are "Manipulate and Search" or "Search and Manipulate". The first of these regimes involves iteratively manipulating sections of data from the comparison signal and assessing its fit to the target data signal. The data is manipulated 'blindly' using random sets of linear corrections applied to the comparison signal. However this method involves large amounts of processing, for very little return. The second regime "Search and Manipulate" minimises the computation needed to perform accurate correction. Using a discrete analysis of the fit between both amplitudes and features embedded in the comparison and target data sets, best fit positions or anchor points are located for each successive shift in the comparison data set. Employing these anchor points, only a single correction is necessary for each successive shift between data signals, dramatically reducing the computational overheads of performing correction tasks. Furthermore, employing a two stage assessment scheme, linear stretches may be applied to successive sub windows of the full signal spectra, allowing accurate and efficient corrections of non linear signal distortions.

4.3.3.1 FSC - Stage One

The first stage of the FSC process employs windows containing from 100% down to 15% of the full complement of samples N contained in the original signal $I_{SET2}(f)$, which are processed in a similar way to that described by Menacer 1992. An assessment of fit between the comparison window and target window is invoked employing the single assessment figure (GDM) of the FSV method, for successive shift positions between windows. Furthermore, the maximum shift allowable within the correction scheme is 20% of the window size under investigation, as shifts greater than 20% introduce severe interpolation errors.

The lowest GDM value obtained from this analysis indicates the position of best fit between the windowed signals under investigation, allowing an assessment of the α term in Equation 4.20 to be obtained. An enhancement to the method of Menacer, employs a discrete analysis of the shifted comparison and target windows, along with the instantaneous *GDM(f)* validation response, allowing single anchor points to be found for each of the compared windows. The lowest value of the *GDM(f)* validation response indicating the best instantaneous sample fit. This step removes the problem of inequality from Equation 4.21, allowing a two part interpolation of the comparison window to be realised by indices transformation, such that:

$$I_{BEST2}(f) = I_{SET2}\left(\frac{A_{SET1}}{A_{SET2}} \cdot f\right) \tag{4.22}$$

where

$$0 < f < A_{SET2}$$

$$I_{BEST2}(f) = I_{SET2}\left(\left[\frac{N-f}{N-A_{SET2}}\right]\cdot\left[\frac{A_{SET1}}{A_{SET2}}\right]\cdot f\right) \quad \textbf{(4.23)}$$

where

$$A_{SET2} < f < N$$

$$I_{BEST2}(A_{SET1}) = I_{SET2}(A_{SET2}) \quad \textbf{(4.24)}$$

where $I_{BEST2}(f)$ indicates the corrected signal sub window, A_{SET1} and A_{SET2} denote anchor points or instantaneous best fit points between the shifted windows, and all values bound by braces *( )* denote indices for their respective signals. Furthermore, as the best shift between windows has been found to be α, it holds that:

$$A_{SET2} = A_{SET1} - \alpha \quad \textbf{(4.25)}$$

where a negative α value denotes a delayed comparison signal , whilst a positive α value indicates an advanced comparison signal with respect to the target. A detailed description of the double interpolation procedure derived in Equations 4.22 - 4.25 is illustrated in Example 4.3.

Example 4.3: Multiple Feature Correction - FSC

Figure 4.23 illustrates the comparison of two signals, exhibiting multiple features embedded in the target and comparison signals respectively. Employing the FSV method and more specifically the GDM for successive shifts between the compared signals, the offset between target and comparison features α is equated to 2 points. Where Figure 4.24 illustrates the best fit shifted version of the comparison and target signals ($\alpha = 2$).

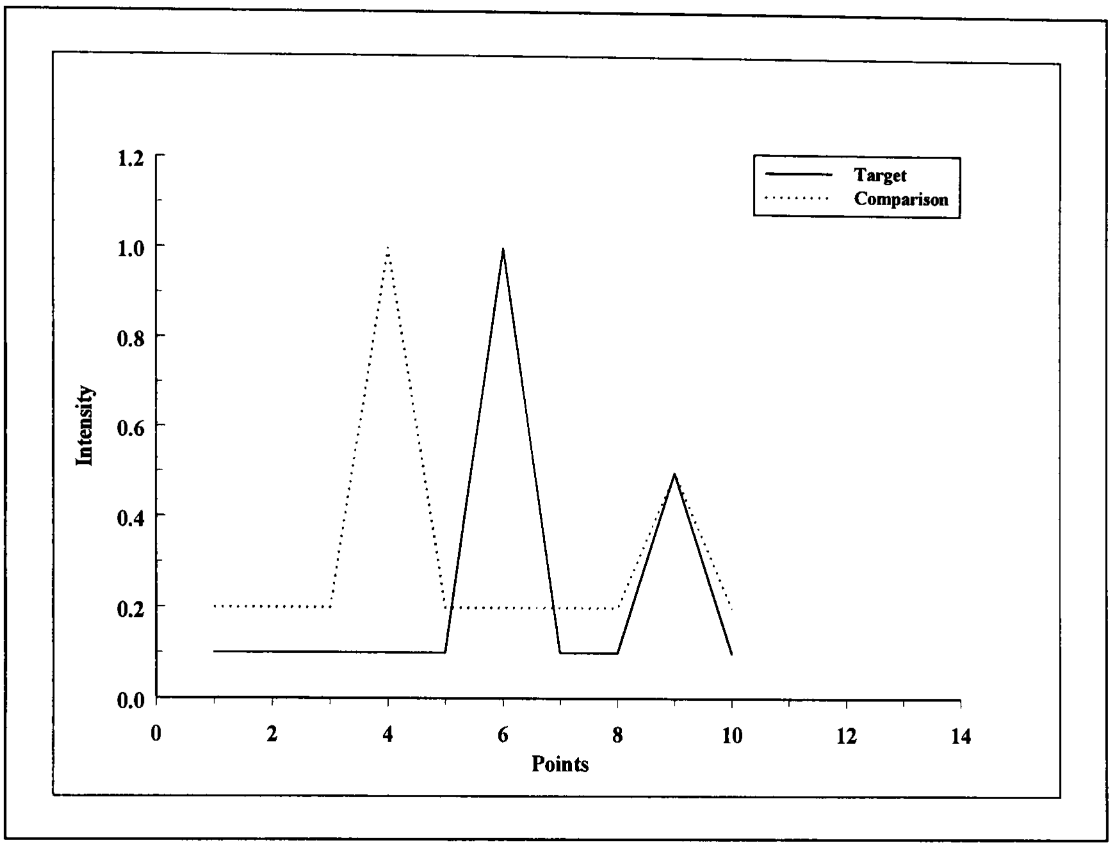

Figure 4.23: Target and comparison signals - multiple features

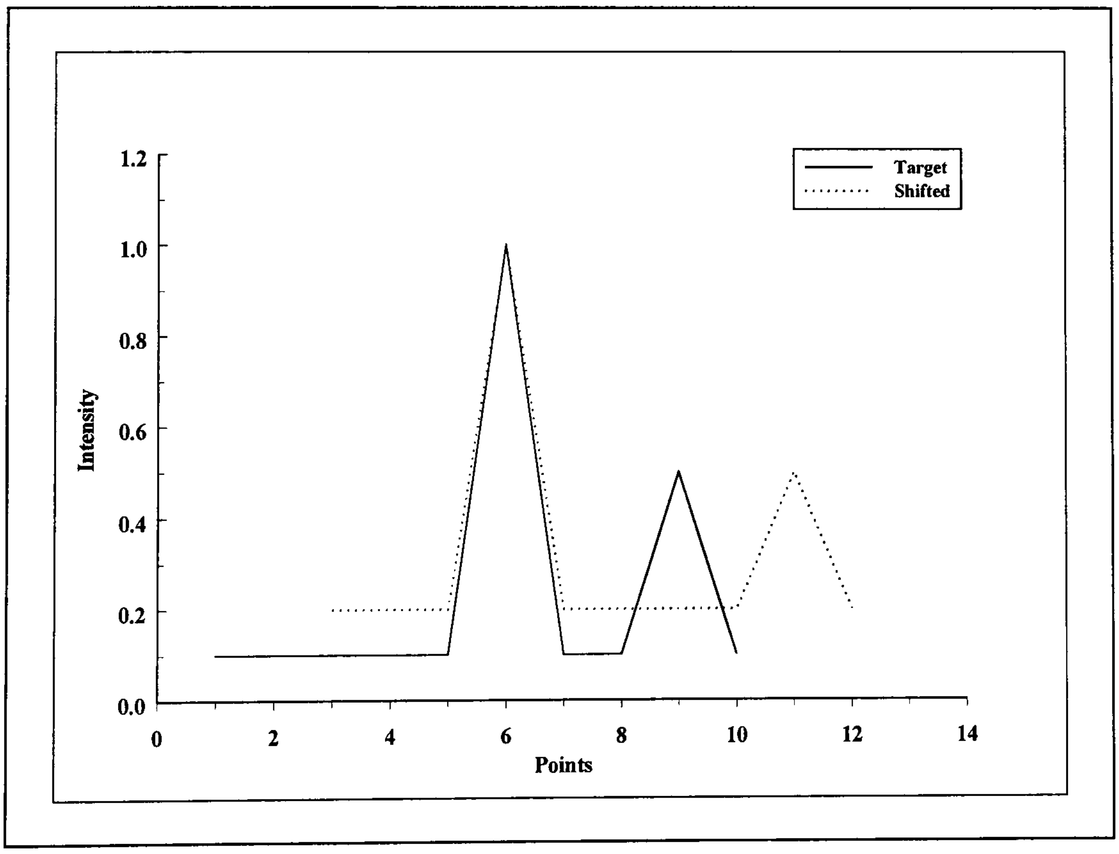

Figure 4.24: Best fit shifted signal - multiple features

Employing the best fit shifted version (Figure 4.24) of the original comparison illustrated in Figure 4.23, a modified instantaneous *GDM(f)* analysis may be applied to locate the best discrete sample fit (anchor points) within each signal. Where, the lowest value β of the modified *GDM(f)* response denotes the best instantaneous fit between the target and comparison signals, whilst the best instantaneous fit for the comparison signal is equated to the lowest value β of the modified *GDM(f)* response minus the best shift position α.

$$Target_{anchor} = A_{SET1} = \beta$$

$$Compaison_{anchor} = A_{SET2} = \beta - \alpha$$

where the modified *GDM(f)* response is analysed at sample positions exhibiting greater than zero first derivative values exclusively. This exclusion analysis is realised through the $FD_{II}(f)$ sub-measure (Equation 4.11) of the *FDM(f)* response employed in the evaluation of the *GDM(f)* response. Modification of the *GDM(f)* response allows an evaluation of best instantaneous fit between samples inherently related to trends and features within the main signal sets. Whilst sections of the signals which do not directly relate to the trends and features being analysed and corrected are disregarded.

The realisation of single anchor points denoted by their respective sample indices for both target and comparison signals, along with the evaluation of the best fit signal shift allows the removal of inequality from Equation 4.21, whilst reducing the computation required to correct distorted data signals. The method of Menacer and more specifically Equation 4.20 underestimates the total stretch required to align two distorted features, and corrective stretches must be applied to the comparison signal iteratively until the cross correlation coefficient reduces to a predefined value. Within the FSC method, a single stretch may be applied in two successive stages employing the anchor points and the best shift position between signals. Furthermore, employing this analysis the full compliment of samples used to construct the comparison signal may be retained without data loss from the signal boundaries.

Employing the FSC method and the shifted data signals illustrated in Figure 4.24, the target and comparison anchor points are 6 and 4 respectively, derived employing the modified *GDM(f)* response.

$$Target_{anchor} = \beta = 6$$

$$Compaison_{anchor} = \beta - \alpha = 6 - 2 = 4$$

where the comparison signal is advanced denoted by a positive α value.

Substituting the target and comparison anchor points along with the best shift value α into Equations 4.22 - 4.24, whilst employing the unshifted target and comparison signals illustrated in Figure 4.23, allows for a single correction of the distorted comparison signal. Furthermore, Equation 4.22 is employed to expand the first section (f_{min} to (β-α)) of the comparison signal to its corrected length (f_{min} to β). Equation 4.23 compresses the second section of the comparison signal ((β-α) to f_{max}) to its corrected length (β to f_{max}). Whilst Equation 4.24 substitutes the comparison sample located at β for its corrected sample located at β-α. The corrected signals are illustrated in Figure 4.25, indicating the expanded and compressed areas of the comparison signal.

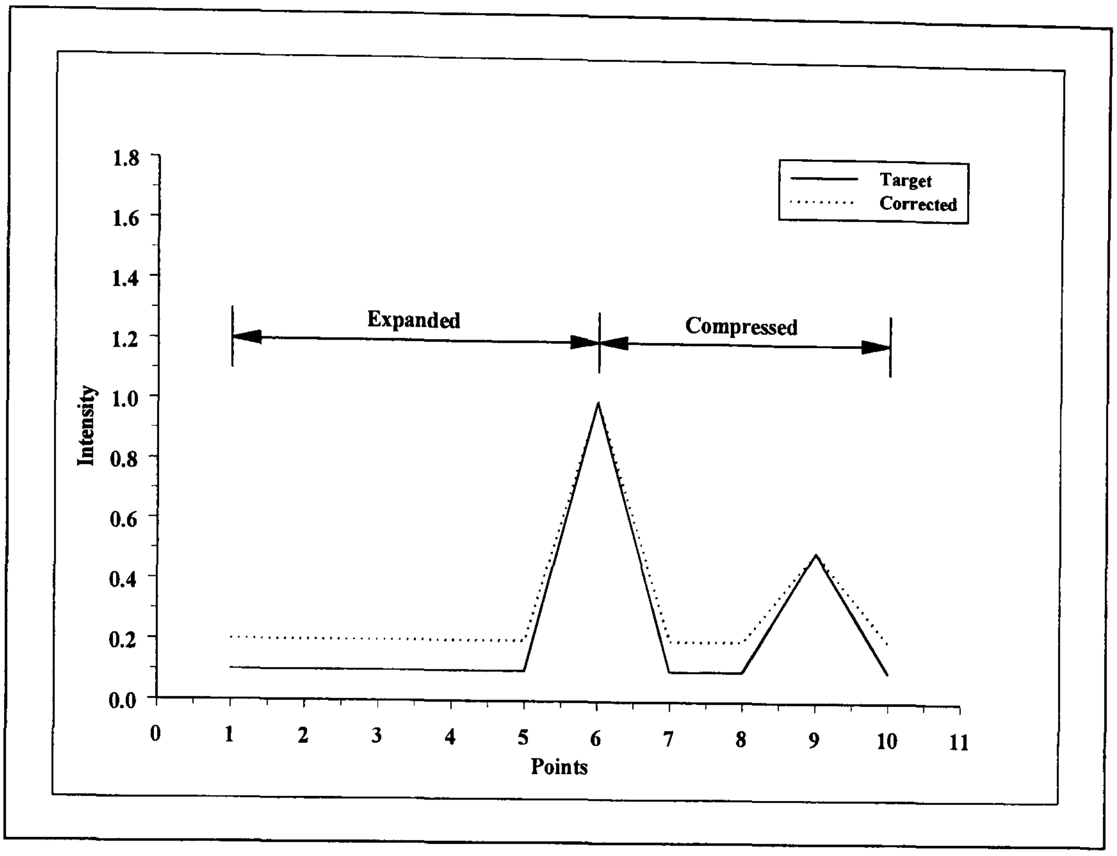

Figure 4.25: Corrected signals employing FSC method

This analysis allows large distortions to be approximated and corrected during the early stages of the analysis, with very little computational overhead.

4.3.3.2 FSC - Stage Two

The second stage of the correction analysis employs windows containing between 15% and 10% of the full complement of samples contained in the distorted comparison signal. Within this stage of the method, considerable accuracy is contained in the corrective analysis applied to the distorted signal. Equations 4.24 - 4.25 are again employed, however, all shift positions are analysed using their respective anchor points. An initial evaluation of α is not obtained employing the lowest GDM value for successive shifts between the comparison and target signals. Hence, each successive shift between the comparison and target window is analysed and stretched employing the modified *GDM(f)* function, as in most cases, the best fit shift position between signals does not necessarily hold the best stretched position for the distorted signal. In general the complexity of the distorted signals will not allow a true evaluation of the 'best shift' position between signals, as stretching one section of the comparison signal may invariably cause degradation of the comparison and target fit in other areas of the signal spectra.

Figure 4.26 illustrates the corrected comparison of $I_{SET1}(f)$ and $I_{SET2}(f)$ illustrated in Figure 4.21. The comparison of Figure 4.26 illustrates significant reduction in the signal distortions observed in the comparison of Figure 4.21.

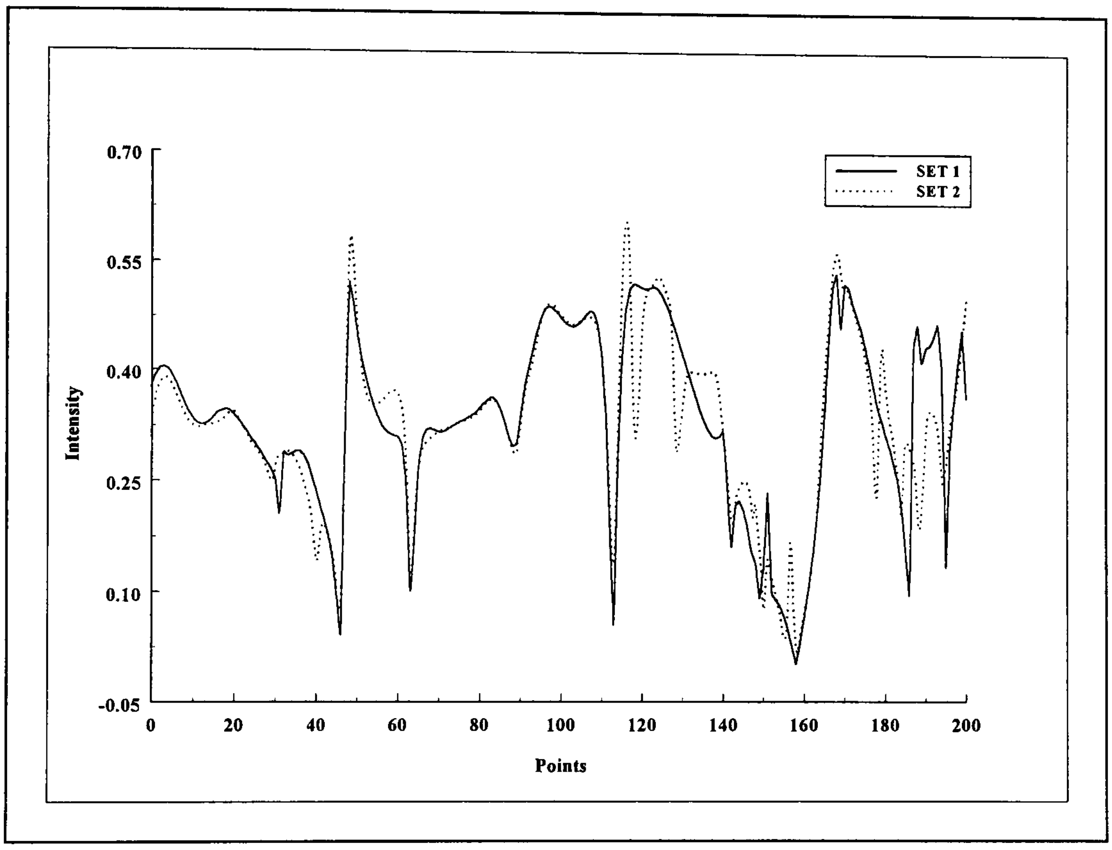

Figure 4.26: Corrected Data Sets - $I_{SET1}(f)/I_{SET2}(f)$

Table 4.3 indicates the uncorrected and corrected FSV results for the comparisons of Figures 4.21 and 4.26 respectively. The corrected results indicate a significant improvement in terms of both amplitude levels and feature shapes and positions. Whilst the GDM indicates an improvement from a 'fair' comparison to a 'good' comparison.

	ADM	***FDM***	***GDM***
Uncorrected	***0.12***	***0.37***	***0.42***
Corrected	***0.09***	***0.24***	***0.27***

Table 4.3: Uncorrected and corrected FSV results - $I_{SET1}(f)/I_{SET2}(f)$

The inclusion of the FSC method in an assessment of distorted signal sets allows a clear indication of the positional and structural discrepancies between compared signals. In the Example of Table 4.3 (employing the GDM values only), the structural (i.e. amplitude and feature) discrepancies are equated to the corrected value of the GDM (0.27 or good). While the positional discrepancy incurred in acquiring the results is equated to the difference between the uncorrected and corrected GDM values (0.42 - 0.27 = 0.15 or very good - good).

4.3.4 Distorted Signal Correction Employing Parallel Processing

Parallel processing, the method of having many small tasks solve one large problem, has emerged as a key enabling technology in modern computing. The past several years have witnessed an ever-increasing acceptance and adoption of parallel processing, both for high-performance scientific computing and for more 'general purpose' applications. This was a result of the demand for higher performance, lower costs, and sustained productivity[Geist 1994].

Parallel processing offers a method of splitting large computational problems into many small tasks, whilst the hardware[Beowulf 1999] (networked computational platforms) and software[Geist 1994, PVM 1999] (Parallel Virtual Machine (PVM)) with which this technology is associated allows these small tasks to be solved concurrently.

Within the parallel processed FSC analysis of distorted signals, each sub-window (illustrated in Figure 4.27) is analysed and corrected in parallel, for each level containing 2 or more sub-windows (i.e. N/2, N/3, etc.). However, whilst the adoption of this technology substantially increases the speed at which distorted signals may be corrected, the implementation of this technology within the FSC method requires several levels of optimisation.

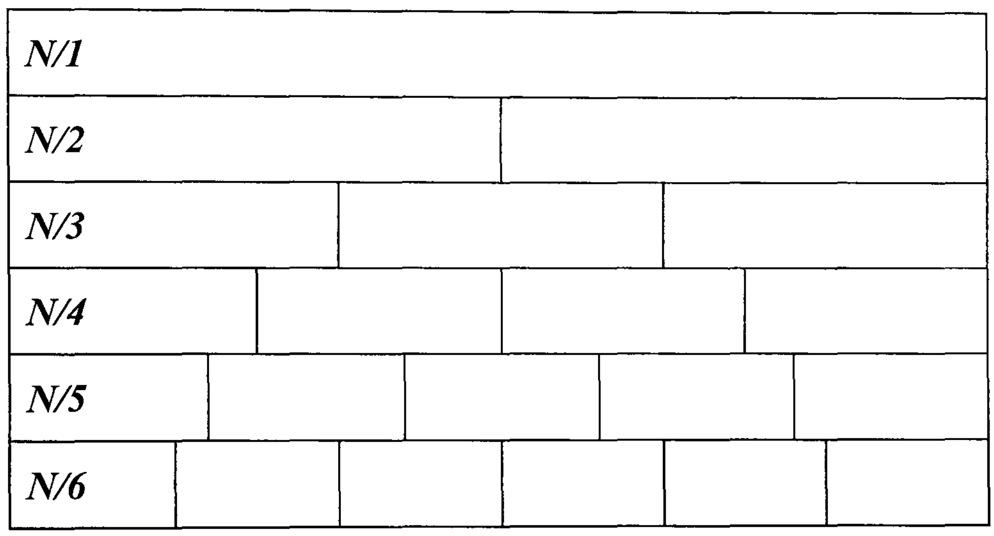

Figure 4.27: Reducing windows scheme employed by the FSC method

Table 4.4 indicates FSC execution speeds using the parallel processing scheme described previously, with signals containing 500 samples, and a parallel virtual machine employing ALPHA 533MHz processors[Alpha 1999] connected through a local 100Mbs/s Ethernet network. The results indicated in Table 4.4 illustrate the potential advantages of employing parallel processing within the FSC methodology. However, it is perceived that substantial improvements may be obtained through the employment of dynamic or adaptive schemes, which manage the processing of parallel tasks(detailed in Section 8.2).

Number of Processors	***Speed of FSC Execution***
1	*≈ 21 Seconds*
2	*≈ 16 Seconds*
3	*≈ 09 Seconds*

Table 4.4: Parallel processing performance (FSC)

4.4 HIERARCHICAL PROCEEDURE - A STRUCTURED ANALYSIS

The hierarchical structure of an assessment employing the FSV and FSC methods are indicated in the seven points below:

1. *FSC method may be applied to correct distorted signals.*
2. *ADT, FDT and GDT values set for subsequent FSV evaluation.*
3. *FSV method applied, GDT, GDM figure of merit and GDM confidence levels employed to assess pass/fail criteria.*
4. *If comparison is of sufficient quality method of data acquisition may be passed.*
5. *If comparison fails, ADM and FDM figures of merit and confidence levels are employed to assess the nature and magnitude of major discrepancies.*
6. *ADM(f), FDM(f) and GDM(f) diagnostic responses employed to locate major discrepancies.*
7. *FSV information used to modify method employed to acquire data sets.*

After an initial evaluation of the compared signals using both the FSC and FSV methods, the GDM information (figures of merit and confidence levels) is employed to evaluate the ***overall*** discrepancies between compared signals. The aim of the GDM is to remove the need for a visual inspection of the compared data, providing information expressing the overall quality of a comparison. If the information obtained from both the GDM figure of merit and GDM confidence levels indicates that the comparison is of sufficient quality (less than the predefined value of the GDT), the data acquisition method may be passed and no further evaluation is necessary.

Conversely, further investigation of the comparison is necessary if the GDM values obtained from the FSV method are greater than the predefined value of the GDT. Further information is obtained from both the ADM and FDM figures of merit and confidence levels, which indicate the nature of major discrepancies impinging on the comparison. After the nature of the errors between compared signals is known, highly detailed diagnostic information in the form of the *GDM(f)*, *ADM(f)* and *FDM(f)* may be employed to assess the locations and magnitudes of discrepancies between compared data sets.

4.5 CHAPTER SUMMARY

Employing the FSC and FSV schemes detailed in this Chapter, methods of data acquisition may be validated and their differences quantified. The quantitative/qualitative values extracted from the FSV method directly relate to the common qualitative classification scale employed by engineers and scientists performing visual evaluations. The fundamental algorithms embedded in the FSV method are based on measures used in a visual comparison of results. Whilst highly detailed diagnostic information is available allowing valid assessments of both the location and magnitudes of errors between compared signals. A user may include a measured level of subjectivity based on either amplitudes or features in a comparison allowing a flexible assessment to be obtained. In this way, results from a wide cross-section of application areas may be validated with the validation information obtained from each being directly comparable (a good comparison in the field of EMC being identical to a good comparison in the field of DNA analysis). Using the information obtained from a comparison, rational decisions can be made on whether further investigation is necessary, or a framework can be established for justifying why additional work is unnecessary. Through the application of the FSV method, a rigorous framework for identifying and disseminating good working practice may be constructed, this is pursued further in Chapters 5 and 6.

CHAPTER 5

COMPARISON OF TECHNIQUES

5. COMPARISON OF TECHNIQUES

The performance of any validation method depends fundamentally on the variability and diversity of the data signals to be validated. This chapter addresses the problem of assessing both the performance and inter-relationship between the automated validation/verification methods detailed in Chapters 3 and 4. The clarification of differences between these automated validation methods is not a simple task, as, in general, the methods employ different quantitative scales. Furthermore, it is not a straightforward procedure to re-scale all methods to common units due to the high complexity embedded in each of the validation algorithms. It is however, relatively straightforward to compare the ability of these validation methods to rank order comparison data sets if compared to a bench mark method, illustrating the relative and not absolute differences between methods.

The combined visual evaluation results of highly skilled subjects possess very high levels of confidence(detailed in Section 2.6). Furthermore, the brain is the best pattern recognition device known and visual evaluation is the most prevalent form of validation in use to date. It is to this end that the combined results of highly skilled subjects performing identical visual evaluations are employed as the bench mark for assessing the performance of automated validation schemes within this Chapter.

This Chapter assesses the performance of four validation schemes namely: correlation, Zanazzi Jona reliability factor, van Hove reliability factor, and the Feature Selective Validation method. Results obtained employing these automated validation techniques are compared with visual evaluation bench mark results. Section 5.1 introduces seven complex data signal comparisons employed to assess the performance of the four automated validation techniques. Section 5.2 presents the results of a survey of subjects involved in visually evaluating the comparisons illustrated in Section 5.1, along with the combined confidence levels associated with each comparison. Section 5.3 assesses the ability of the four automated validation techniques to rank order comparisons of complex data signals, along with a summary of features embedded in each of the validation techniques. Further discussions summarise the operation and performance issues related to the four validation methods under investigation.

5.1 COMPARISON DATA SIGNALS

In order to compare the automated validation methods of Chapters 3 and 4 namely: correlation, Zanazzi Jona, van Hove, and FSV, seven data signal comparisons were employed. The seven comparison data sets were chosen in an attempt to gain potentially different comparisons of complex data. All comparisons employ abscissa units of points (samples employed to represent the signals) except comparison three which employs an abscissa scale of frequency. The seven comparisons are illustrated in Figures 5.1 - 5.7.

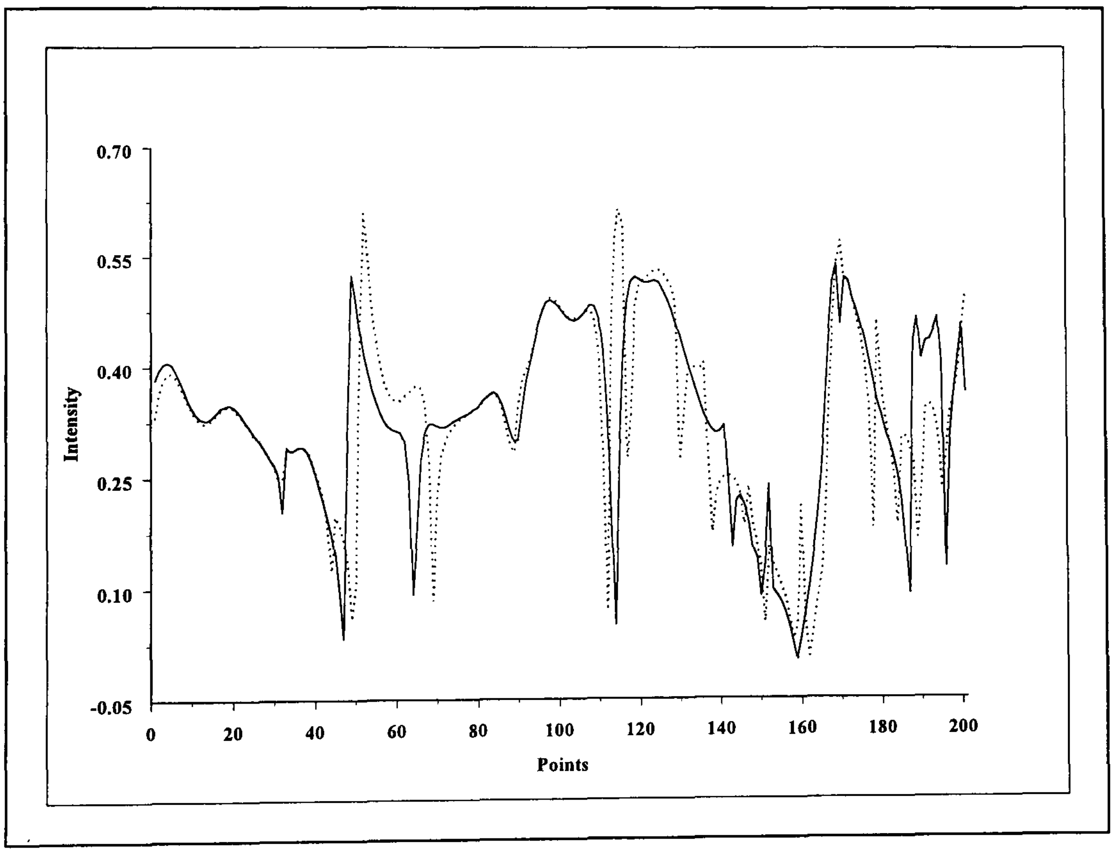

Figure 5.1: Comparison 1

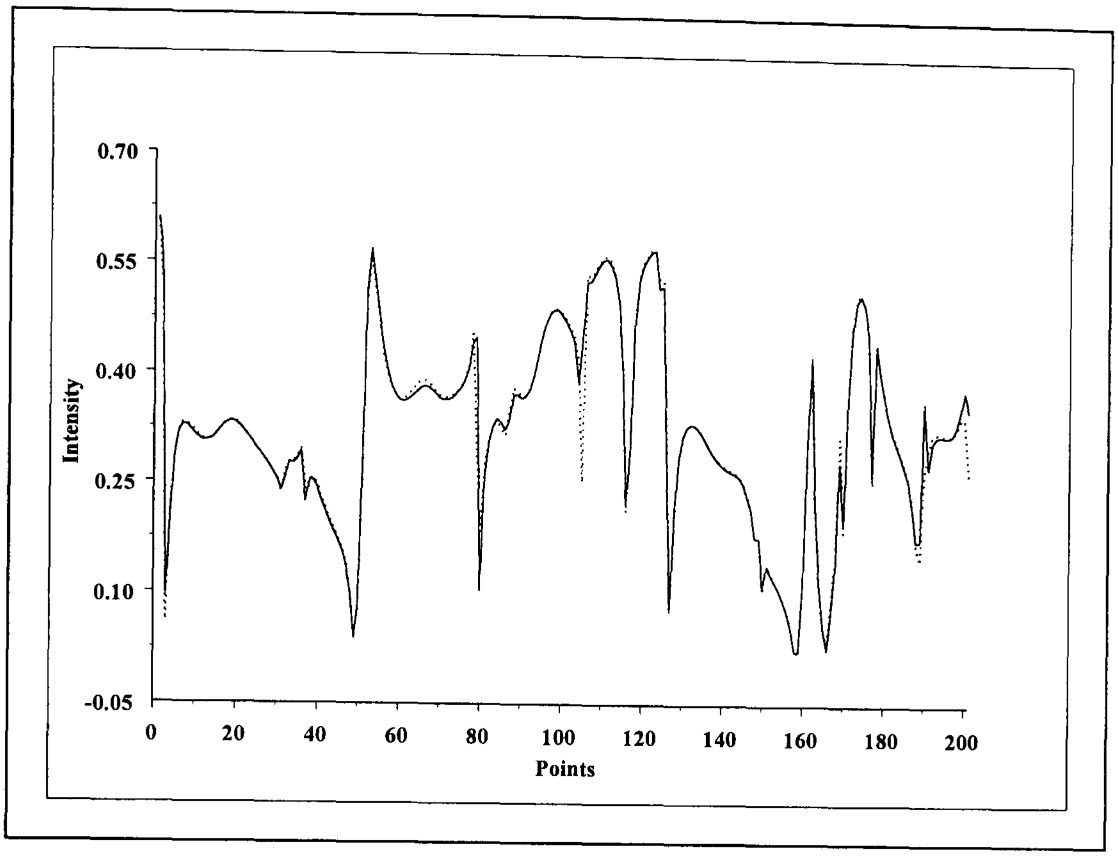

Figure 5.2: Comparison 2

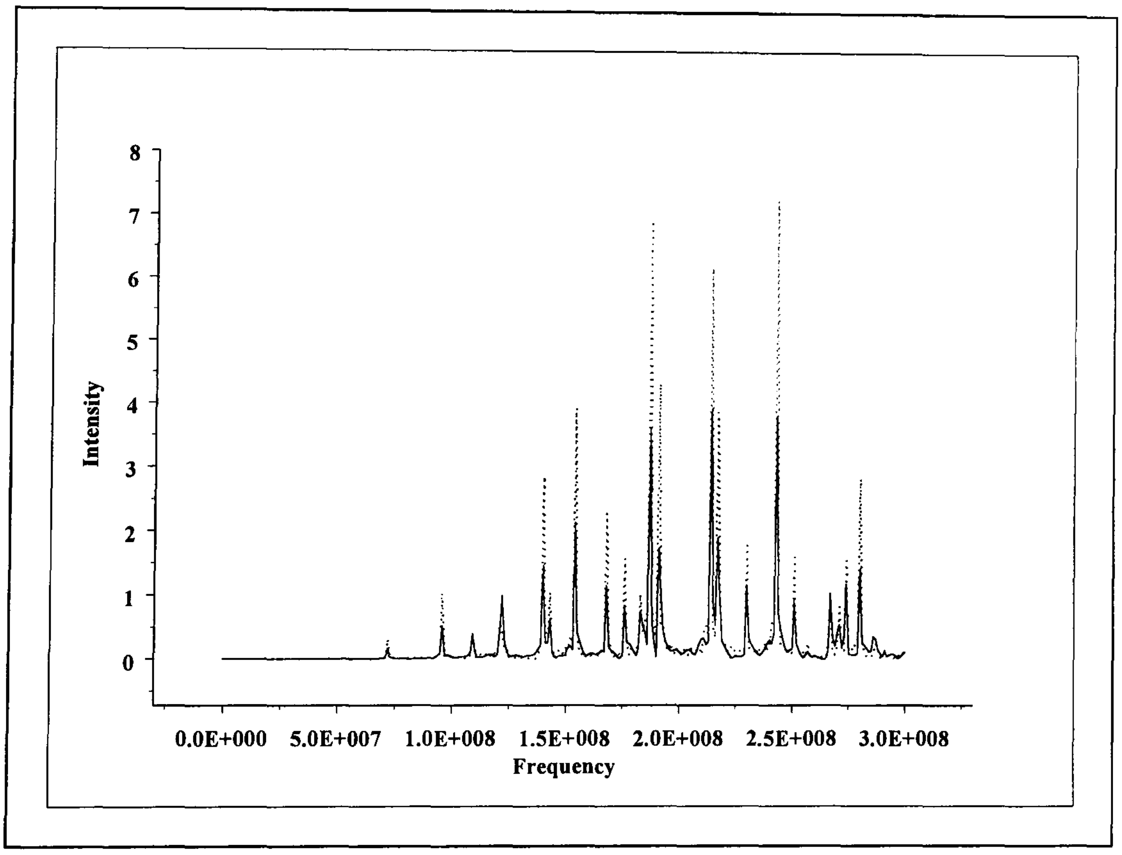

Figure 5.3: Comparison 3

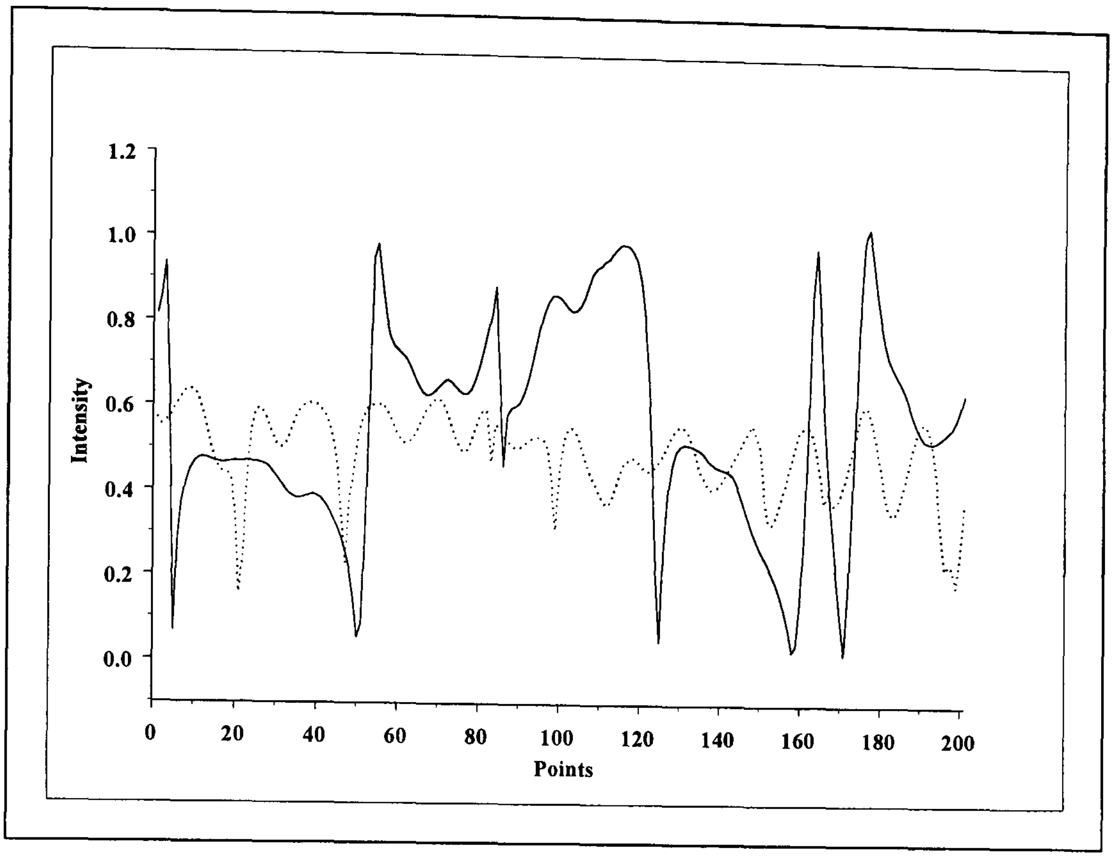

Figure 5.4: Comparison 4

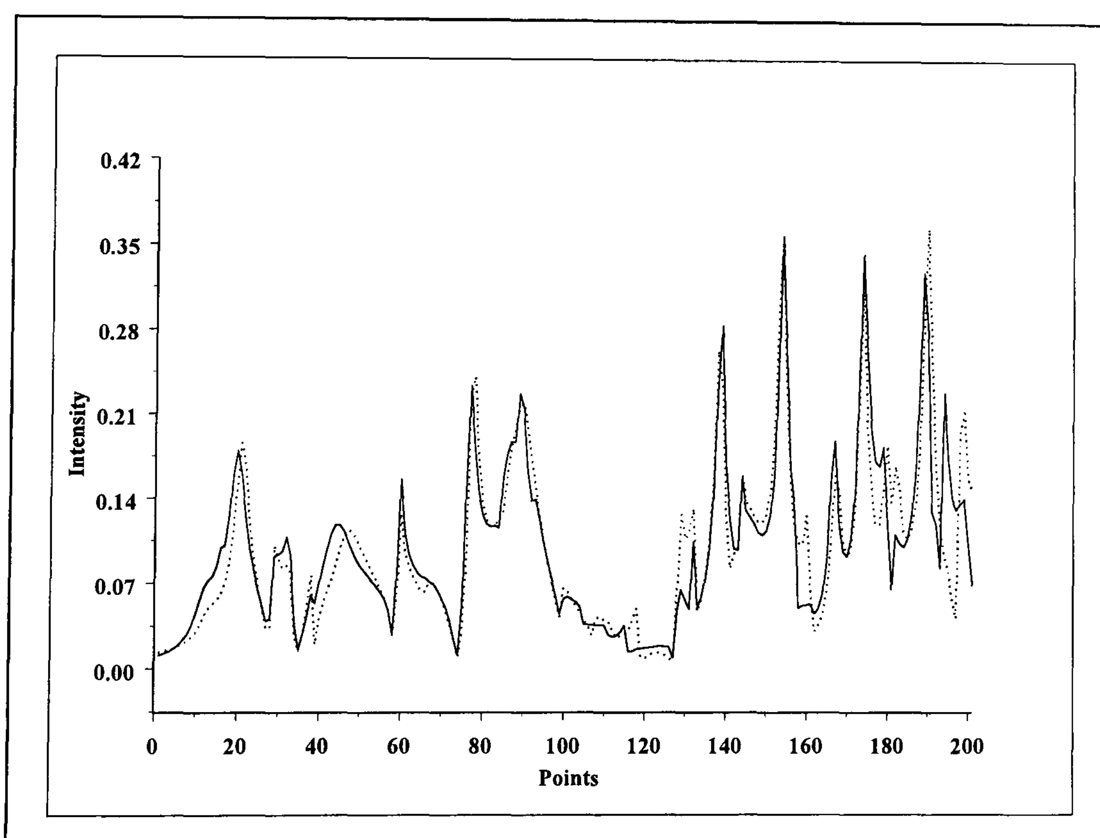

Figure 5.5: Comparison 5

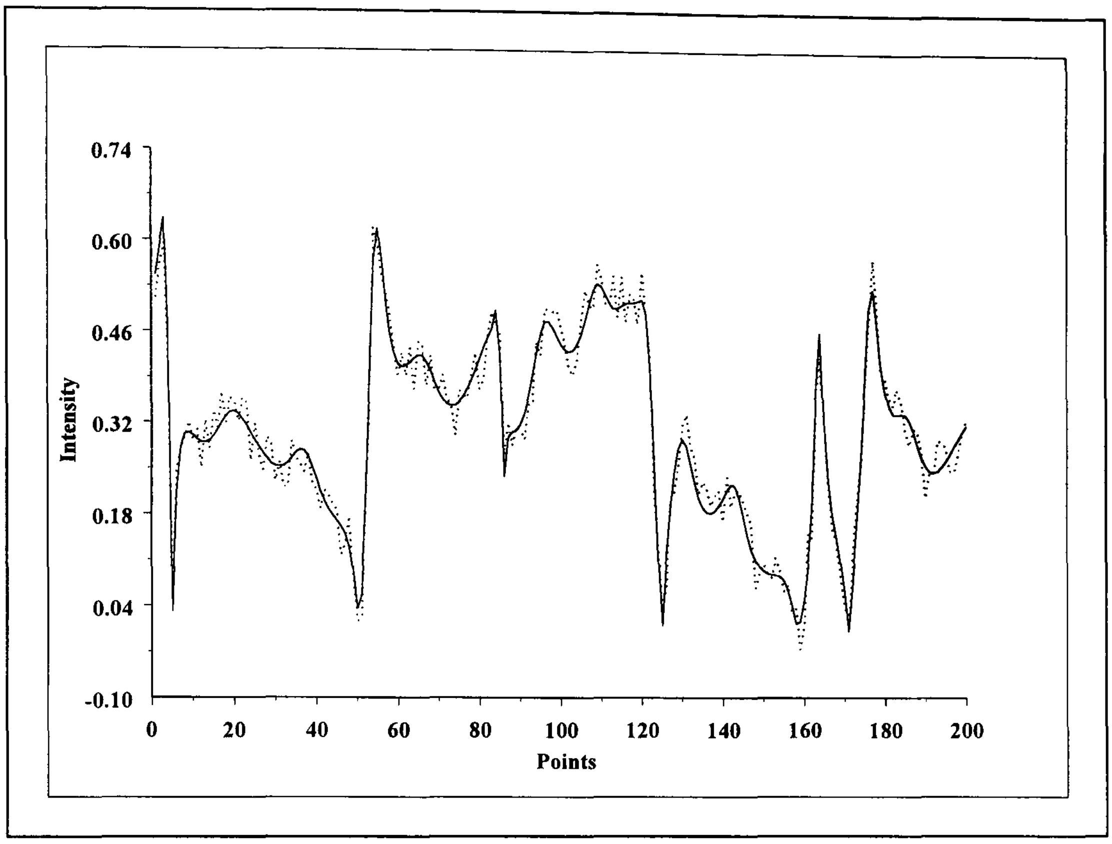

Figure 5.6: Comparison 6

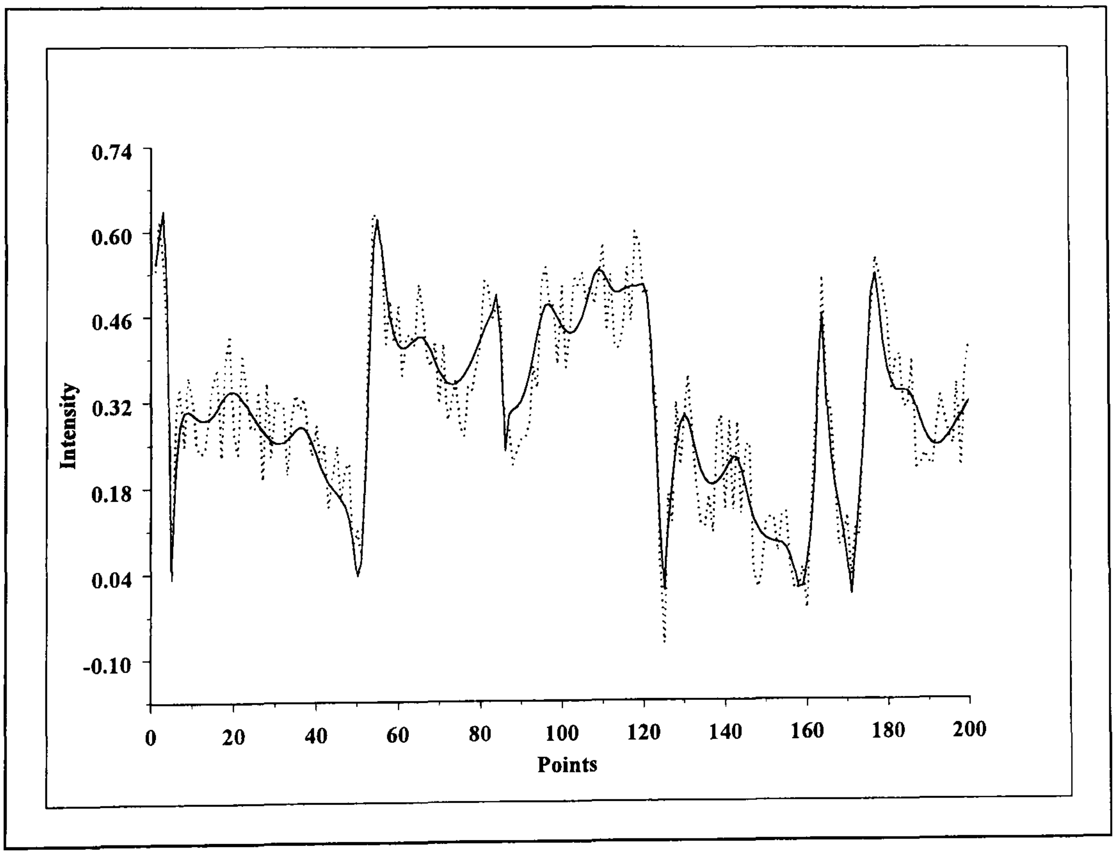

Figure 5.7: Comparison 7

5.2 VISUAL EVALUATION - BENCH MARK RESULTS

Twenty subjects participated in the experiment. The task required each subject to visually assess the seven comparisons illustrated in Figures 5.1 - 5.7, associating each with one of seven quality bands or categories, namely: 'ideal', 'excellent', 'very good', 'good', 'fair', 'poor' or 'extremely poor'. Specific information on the general procedure employed in acquiring each of the comparison sets was not specified. Examples of the experiments were not included in the general task information and no explanation of the meaning of each category was specified. Furthermore, all subjects participating in the experiment were trained engineers and scientists in an attempt to minimise variability between assessment results.

The results from this study - illustrated in Figures 5.8 - 5.14 - were processed, with each quality band given a value indicating the percentage of subjects selecting that category to express the quality of the comparison in question.

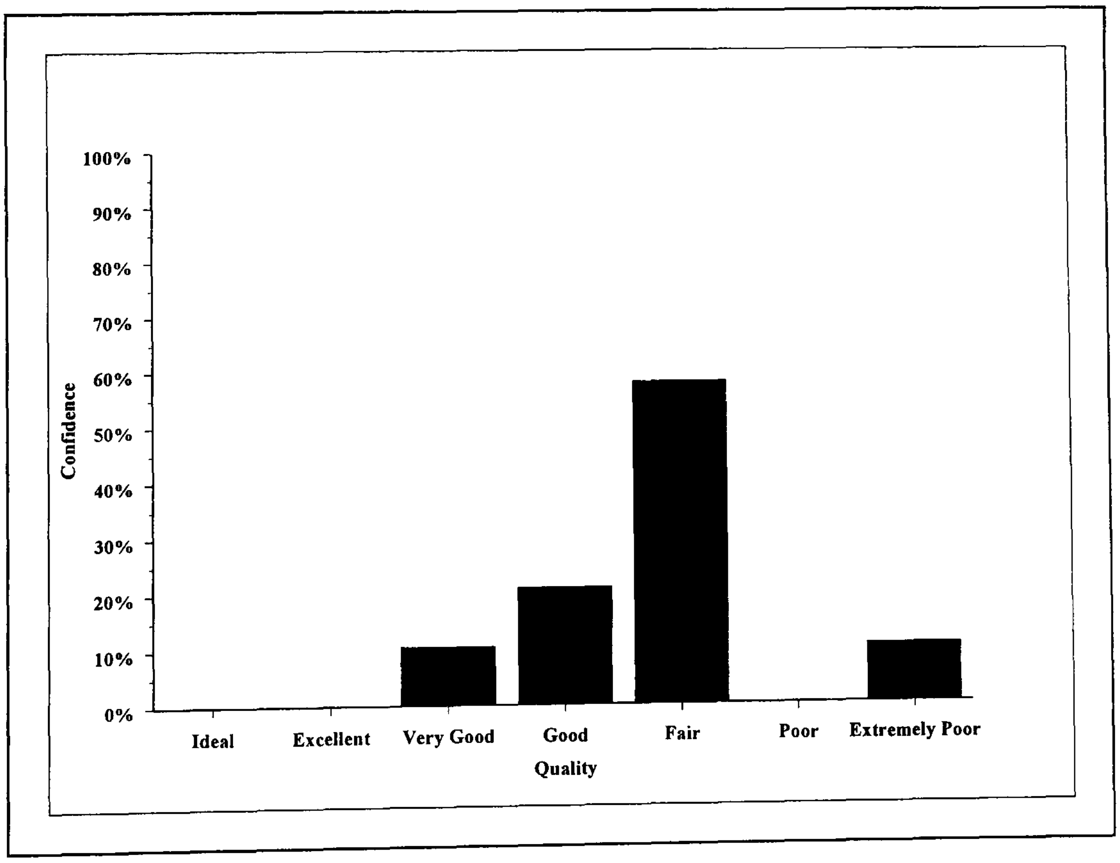

Figure 5.8: Visual evaluation results - comparison 1

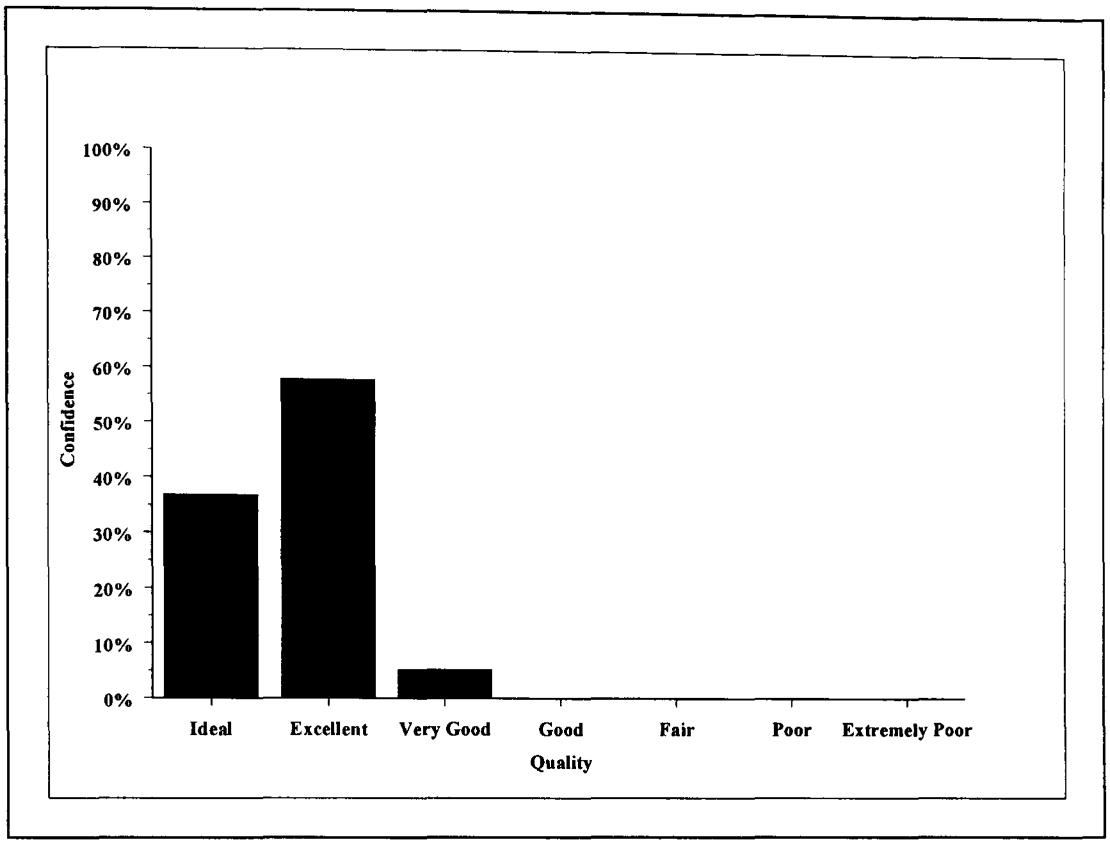

Figure 5.9: Visual evaluation results - comparison 2

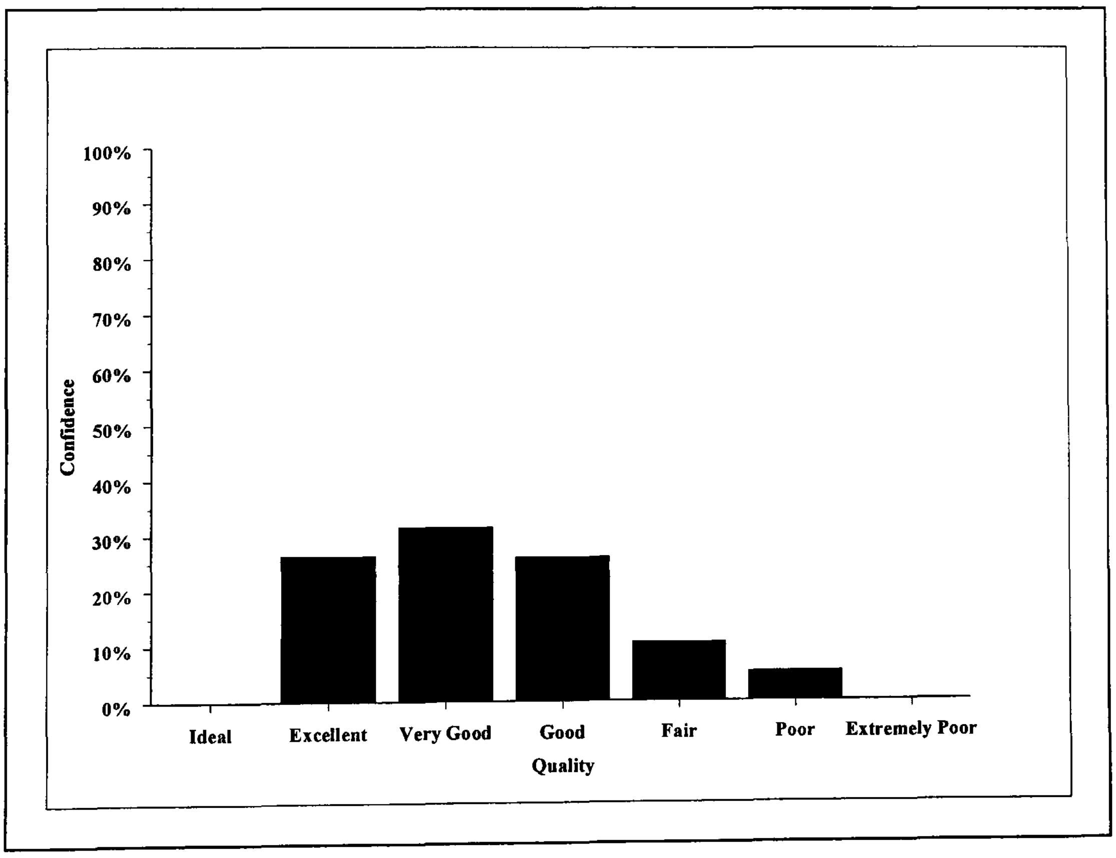

Figure 5.10: Visual evaluation results - comparison 3

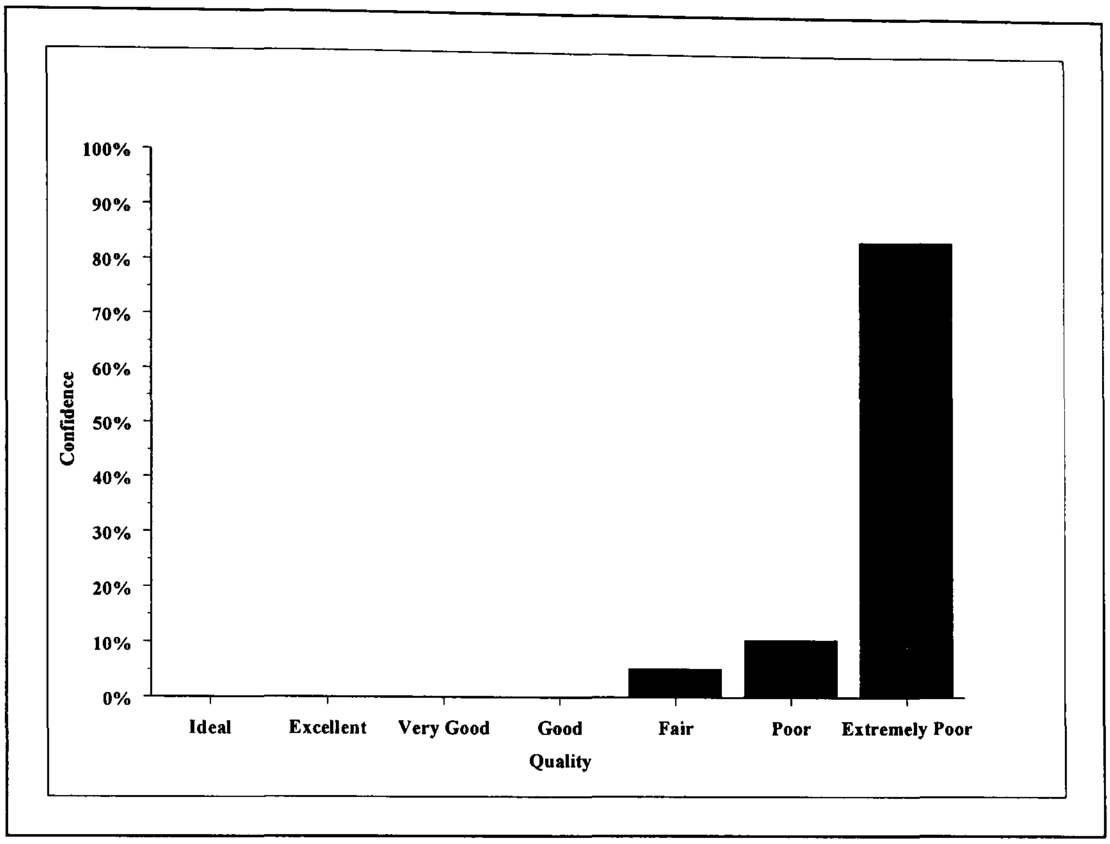

Figure 5.11:Visual evaluation results - comparison 4

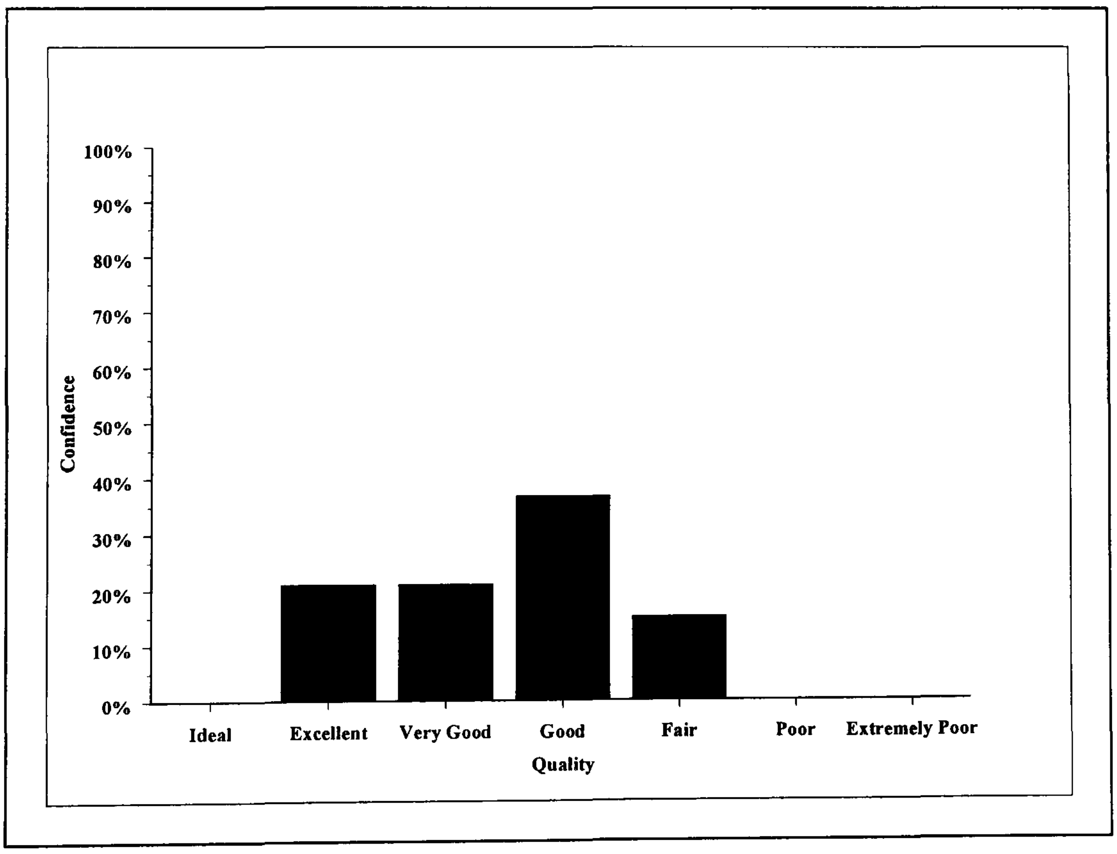

Figure 5.12: Visual evaluation results - comparison 5

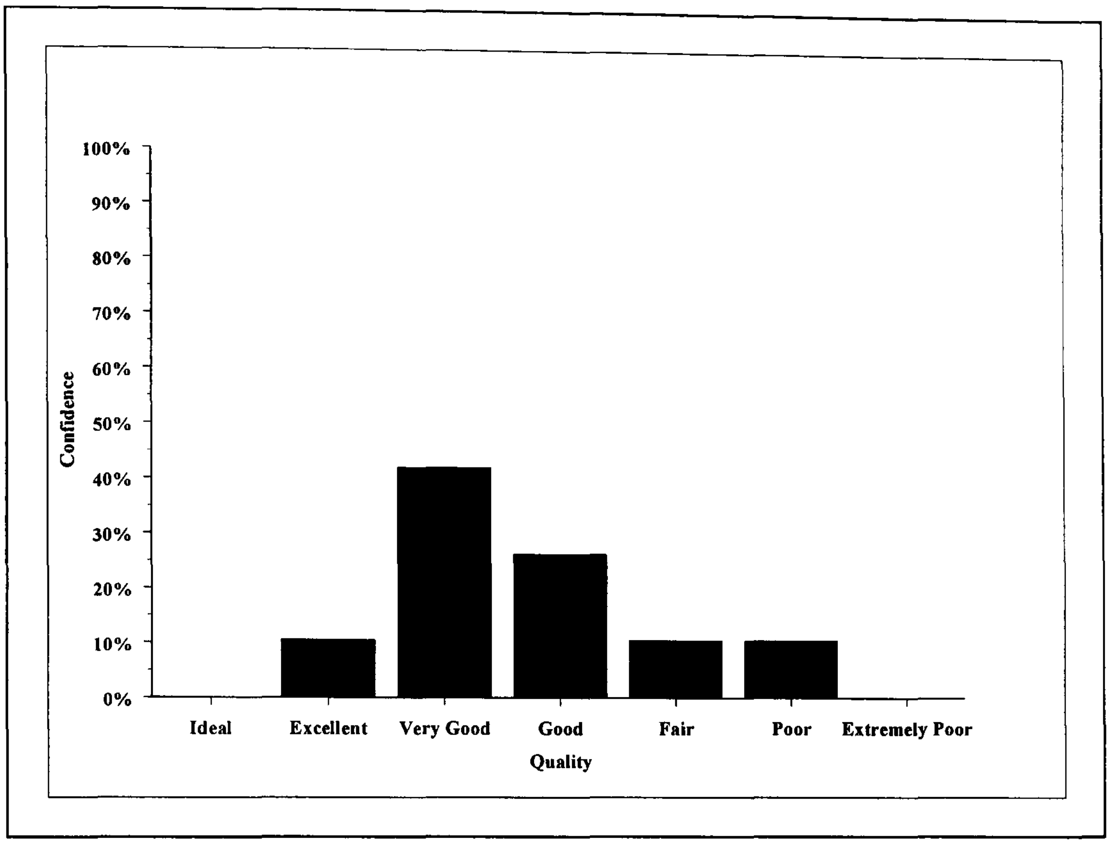

Figure 5.13: Visual evaluation results - comparison 6

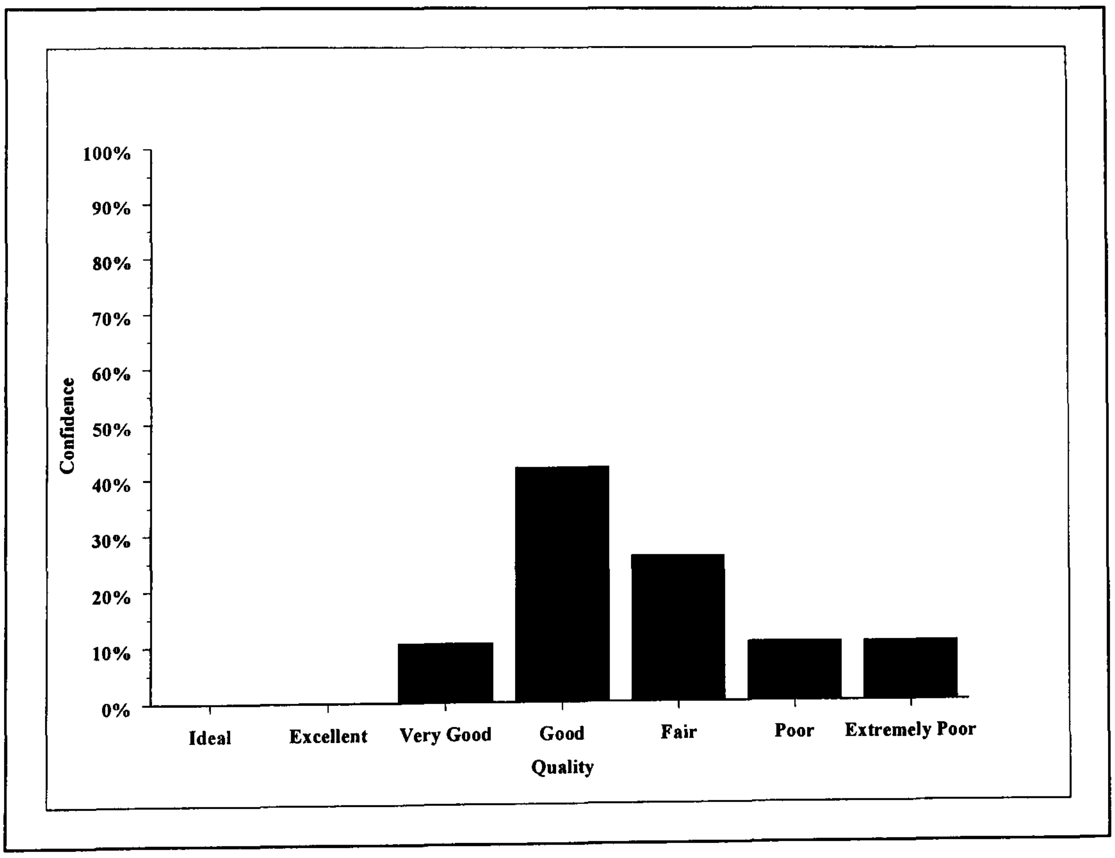

Figure 5.14: Visual evaluation results - comparison 7

The results illustrated in Figures 5.8 - 5.14 indicate the combined confidence levels associated with the seven comparisons illustrated in Figures 5.1 - 5.7 along with the inherent assessment variability between subjects participating in the study. The results of Figures 5.8, 5.9 and 5.11 indicate very little variability between the assessments of subjects visually evaluating the quality of the compared data signals in question. This is indicated by a single, isolated and dominant category effect (major category) and generally low deviation from this dominant category effect. This is more pronounced in the results of Figure 5.11. The results of Figures 5.10, 5.12, 5.13 and 5.14 do not however indicate such clear category effects, as in general, the major categories are not sufficiently pronounced to indicate extremely high levels of confidence. However, the results illustrated in Figures 5.8 - 5.14 all indicate single major category confidence levels and a measured level of confidence may be associated with each comparison set. It should be noted that, in general, comparison results possessing low levels of variability (narrow category effects) exhibit extremely high levels of confidence.

5.2.1 Visual Evaluation Summary

Figures 5.8 - 5.14 illustrate the results of human categorisation effects whilst visually evaluating the seven independent comparisons of Figures 5.1 - 5.7. The results illustrate considerable human variability within the twenty trained engineers and scientists that undertook the seven visual evaluations of complex data. However, within each case a single category (***major category***) possesses a substantial quantity of the overall confidence.

All comparisons indicate a clear level of confidence which may be associated with each compared data signal set. Wide variations in the choice of quality bands is observed in most of the comparisons, indicating considerable human variability. However, single category confidence bandwidths may be associated with each of the comparisons, indicating a measured level of confidence for each comparison.

Table 5.1 indicates the rank ordered results of Figures 5.8 - 5.14 in terms of the level of confidence associated with the major category derived for each comparison set. In the case of comparisons 5 and 7 however, this analysis was not sufficient to rank the two comparison sets, where the major categories indicate that comparison 7 is in better agreement than comparison 5. However, close inspection of the surrounding confidence levels do not lend themselves to such a conclusion, as approximately 40% of the total confidence associated with comparison 5 is associated with the two quality bands ('excellent' and 'very good') below the major category ('good'). Conversely 40% of the total confidence associated with comparison 7 is associated with the three quality bands ('fair', 'poor' and 'extremely poor') above the major category ('good'). It is to this end that comparison 5 is ranked in closer agreement than comparison 7.

Comparison (rank ordered)	*Major category (confidence)*	*Major category (quality)*
Comparison 2	*58%*	*Excellent*
Comparison 6	*42%*	*Very Good*
Comparison 3	*31%*	*Very Good*
Comparison 5	*36%*	*Good*
Comparison 7	*42%*	*Good*
Comparison 1	*57%*	*Fair*
Comparison 4	*84%*	*Extremely Poor*

Table 5.1: Rank ordered visual evaluations

5.3 AUTOMATED VALIDATION VERSUS VISUAL EVALUATION

Employing the automated validation methods detailed in Chapters 3 and 4, global figures of merit expressing the quality of each of the seven comparisons illustrated in Figures 5.1 - 5.7 where derived. The results of this study are presented in Table 5.2. It should be noted that all FSV results were gained without the employment of the FSC method and the optional weighting factors described in Sections 4.3 and 4.2.2 respectively.

	Quantitative values						
Method	*1*	*2*	*3*	*4*	*5*	*6*	*7*
Correlation	***0.035***	***0.001***	***0.032***	***0.890***	***0.310***	***0.002***	***0.009***
Zanazzi Jona	***0.107***	***0.005***	***0.035***	***0.116***	***0.006***	***0.008***	***0.032***
van Hove I	***2.959***	***0.221***	***0.652***	***1.827***	***0.822***	***0.627***	***1.419***
van Hove II	***0.630***	***0.091***	***0.174***	***0.486***	***0.180***	***0.110***	***0.298***
FSV	***0.412***	***0.035***	***0.199***	***0.860***	***0.243***	***0.156***	***0.276***

Table 5.2: Quantitative automated validation results

Table 5.3 illustrates the rank ordered results of Table 5.2, along with the visual evaluation results of Table 5.1.

Method	*Rank order - best to worst (comparison)*						
Visual	*2*	*6*	*3*	*5*	*7*	*1*	*4*
Correlation	*2*	*6*	*7*	*5*	*3*	*1*	*4*
Zanazzi Jona	*2*	*5*	*6*	*7*	*3*	*1*	*4*
van Hove I & II	*2*	*6*	*3*	*5*	*7*	*4*	*1*
FSV	*2*	*6*	*3*	*5*	*7*	*1*	*4*

Table 5.3: Qualitative automated validation results

Table 5.4 compares the performance of the four automated validation methods previously defined in terms of their ability to accurately rank order comparisons compared to the combined visual evaluation results of highly skilled subjects.

Furthermore Table 5.4 implements, a comparison of embedded features within each of the automated validation methods.

	Validation method				
Properties	***Correlation***	***Zanazzi Jona***	***van Hove***	***FSV***	***Visual***
Ψ ***Performance***	***71.43%***	***42.86%***	***71.43%***	***100%***	***Bench mark***
Global measure	√	√	√	√	√
Diagnostics			*	√	√
Weightings			*	√	√
Scaled results			×	√	√
Confidence			×	√	√
Correction	***Linear***	×	×	***Complex***	***Complex***

Table 5.4: Validation performance

Ψ ***Ability to rank order the quality of comparisons***

√ ***Implemented***

* ***Implemented in modified method***

× ***Not implemented in current method***

5.4 SUMMARY OF METHODS

5.4.1 Correlation

The method of correlation indicates a measure of best fit for successive shift positions between data sets. Great care must be taken with the employment of this analysis, as in general, a measure of best fit is not evaluated. A measure of best fit for a section (and not the full compliment) of data is, in reality, evaluated. However, a measure of validity must be placed on the full compliment of results and this may be achieved by simple evaluation employing a correlation algorithm along with non shifted signals (i.e. $R_{12}(0)$ - see Examples 3.1 and 3.2).

The quantitative values gained from the method of correlation indicate the global similarity to a maximum of one for the comparison in question. Correlation employs no weighting structure, and treats all parts of the compared signals as equal; there is no latent weighting for amplitudes, or features. A further drawback within the method of correlation stems from the inherent nature of the algorithm from which results are obtained. Correlation employs a measure of similarity derived from instantaneous multiplication methods, this type of analysis does not lend itself to the derivation of a point by point or diagnostic assessment of compared signals (detailed in Section 3.1). This in turn renders the method powerless in cases where high level diagnostic information is required.

Enhancements to the method of correlation[Menacer 1997] detailed in Section 4.3 have illustrated the method's ability to correct linearly distorted signals. However, the inability to provide powerful diagnostic information based on isolated sections of these distorted signals inhibits the correction methods ability to correct non linear distortions within compared signals. Furthermore, the equations employed to correct distortions discard data from the ends of the signals, this in turn forces the equations to act on the full spectrum of data within the signals only, as data loss within the boundaries of the full spectrum would induce significant errors in the corrected data signals (see Example 4.2).

5.4.2 Zanazzi Jona

Results obtained employing the Zanazzi Jona reliability factor provide powerful global information on the reliability of features between compared signals. First and second order derivative differences are applied to extract and emphasise low level features or trends and higher order features respectively. Results gained using the Zanazzi Jona reliability factor illustrate areas of poor reliability between features, whilst providing global information on the overall quality of the comparison. Originally the method of Zanazzi Jona was exclusively employed to gain a global figure of merit expressing the quality of a comparison. However, modifications to the algorithm may allow the collection of discrete validation results allowing diagnostic information to be extracted from a comparison, with very little increase in computational overhead. A major shortcoming in the method of Zanazzi Jona is its inability to provide information about the quality of amplitude levels or general trends between compared signals. This inadequacy to mirror the key measures taken into account during a visual evaluation of results adversely affects the performance of the algorithm in comparison to combined visual evaluation results.

5.4.3 Van Hove

The results of van Hove overcome many of the problems incurred in both the methods of correlation and Zanazzi Jona, by subdividing the difference measurement employed in an analysis of compared signals. These sub-measures, namely amplitude and low level trend differences may be employed independently, or combined to form a global figure of merit, acting on the full compliment of acquired data. However, it should be noted from Section 5.1 that narrow features may have an overriding or masking effect on the true amplitude levels within a comparison, and should be removed during an evaluation of trend and amplitude discrepancies.

Modifications (detailed in Section 3.2.3) to the reliability factor developed by van Hove illustrate the potential advantages of a point by point or diagnostic analyses of complex data signals. However, the measures employed to gain information expressing the quality of a comparison do not mirror those employed by subjects performing visual evaluations. This adversely affects the methods ability to produce information in a categorical manner which is directly related to the combined results of visual evaluations.

5.4.4 Feature Selective Validation

Results obtained employing the FSV method indicate significant improvements over the methods of both correlation and reliability factors in accurately categorising comparisons of complex data signals. The FSV method benefits from the ability to mirror human perception, whilst producing information which is directly related to human variability and the confidence associated with it.

The Amplitude Difference Measure (ADM) included in the FSV method employs low pass filtered data obtained from the signals under investigation, eliminating the potential threat of narrow features masking the validation results. This is illustrated in the FSV results indicated in Table 5.5 obtained employing comparisons 6 and 7. These comparisons employ an identical reference signal $I_{SET1}(f)$, illustrated in Figures 5.6 and 5.7 respectively. However, $I_{SET2}(f)$ included in comparisons 6 and 7 comprise $I_{SET1}(f)$ with an added noise signal. Furthermore, the noise signal added to $I_{SET2}(f)$ in comparison 7 is of greater magnitude in comparison to the noise signal added to $I_{SET2}(f)$ in comparison 6. An automated validation method which boasts minimal errors due to the masking effects of narrow features embedded in a comparison of signals should, as a minimum, exhibit amplitude difference results of similar magnitude for the two comparisons described previously. Furthermore, the major component of discrepancy between these sets of compared signals should be feature shapes/positions.

	ADM	*FDM*	*GDM*
Comparison 6	*0.02*	*0.15*	*0.15*
Comparison 7	*0.05*	*0.26*	*0.27*

Table 5.5: Component FSV measures - comparisons 6 and 7

The component FSV measures indicated in Table 5.5 illustrate the methods ability to accurately assess complex data signals employing two separate measurements. These results illustrate the advantages of isolating homogeneous sections of the compared data signals before discrepancies between amplitude levels and feature shapes/positions are assessed.

The confidence levels included in the FSV method - illustrated in Figures 5.15 - 5.21 (shown against visual evaluation confidence levels) - may be viewed as an extension to the statistical analysis applied during an assessment of compared signals. This analysis is unique among modern automated validation schemes and is introduced to mirror the confidence levels derived from the combined visual evaluation results of highly skilled engineers and scientists. Quantitative information regarding the variability involved in visually assessing compared data is also embedded in each of the FSV confidence plots. As, in general, a direct relationship is observed between the percentage confidence associated with a comparison and the spread or distribution of the confidence levels. Consequently, comparisons exhibiting highly distributed confidence levels will in general possess broader confidence bandwidths (less confidence per quality band) than comparisons exhibiting less distributed confidence levels. This phenomenon is illustrated in Figure 5.18 and Figure 5.17 respectively.

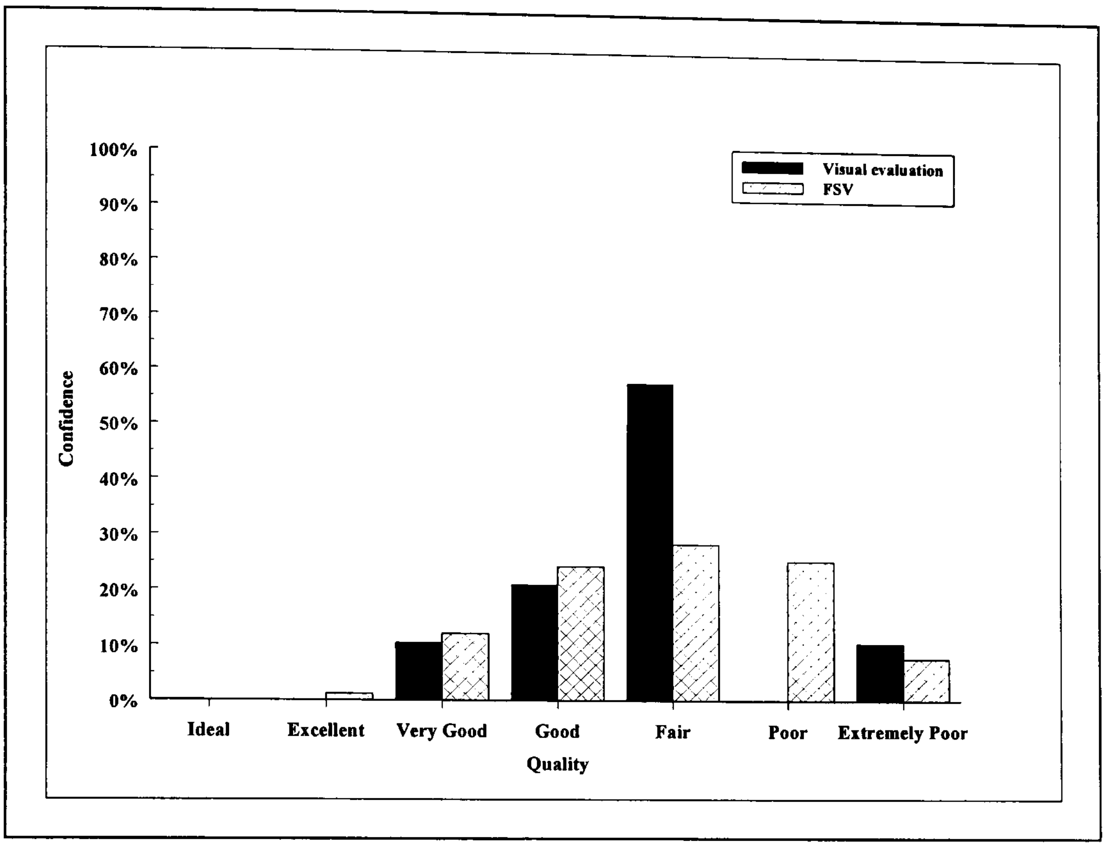

Figure 5.15: Visual / FSV confidence levels - comparison 1

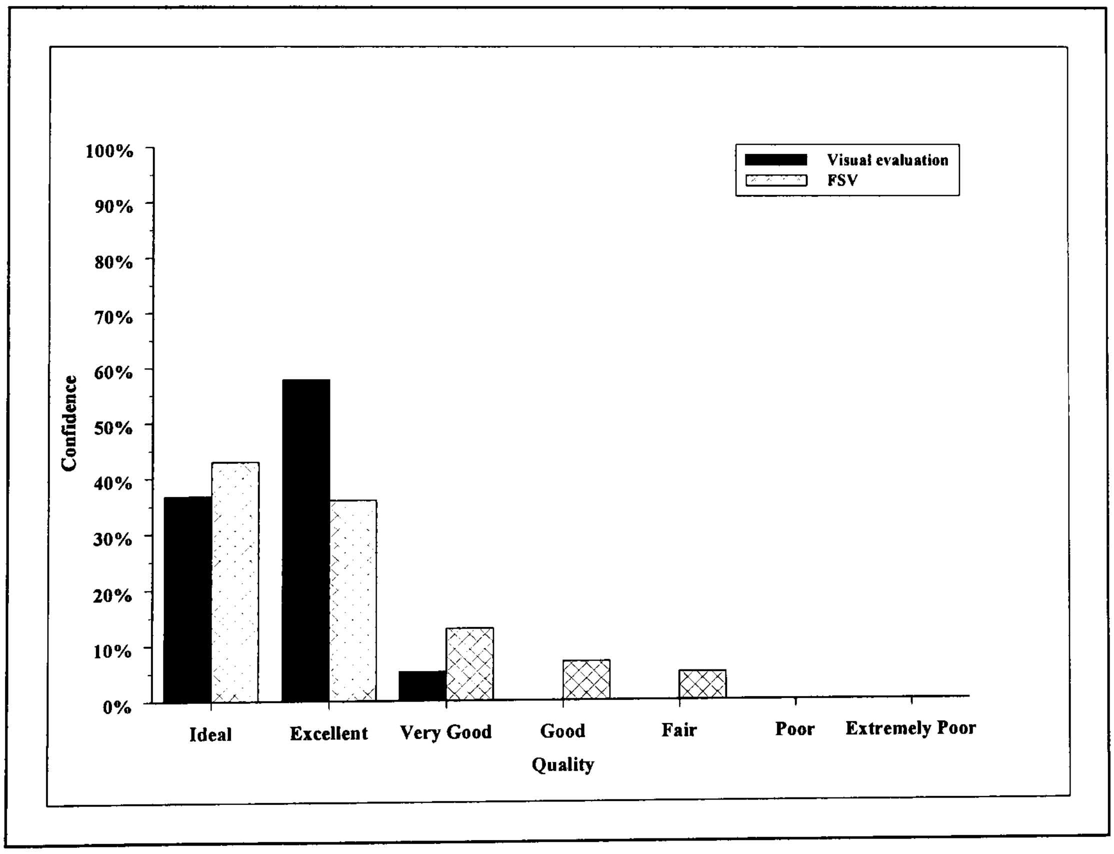

Figure 5.16: Visual / FSV confidence levels - comparison 2

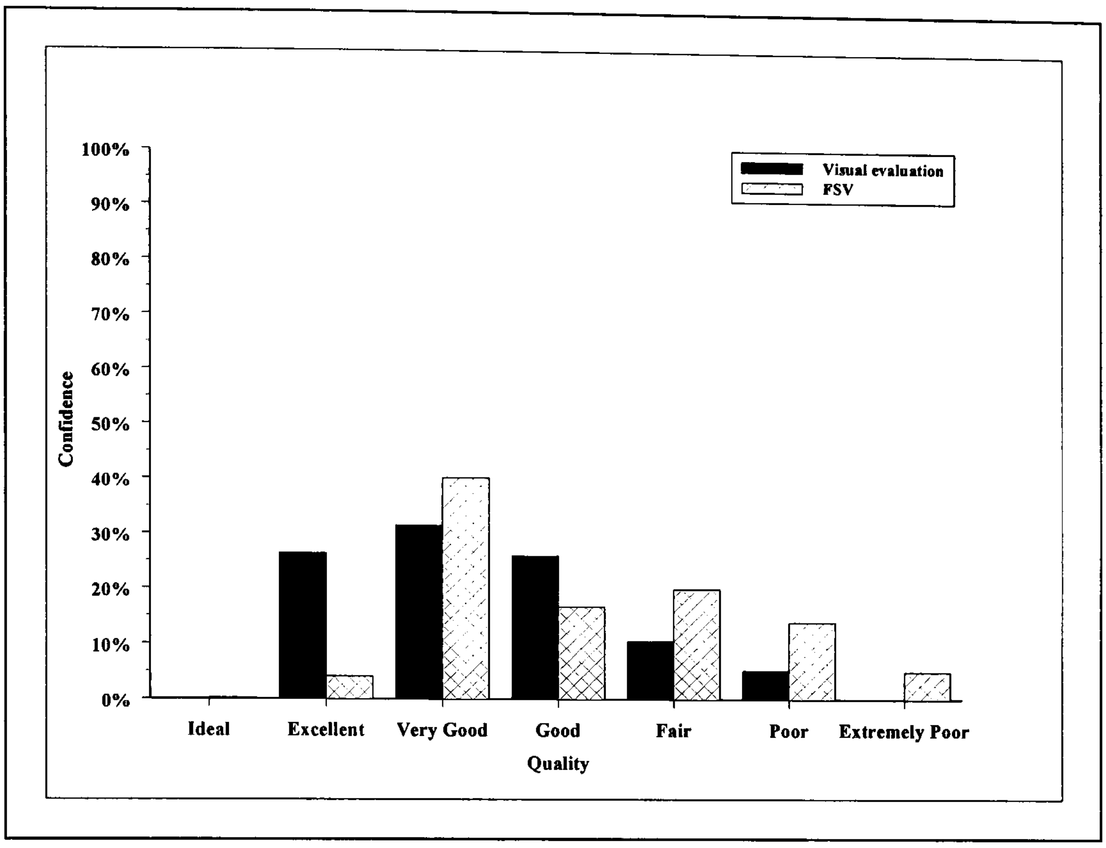

Figure 5.17: Visual / FSV confidence levels - comparison 3

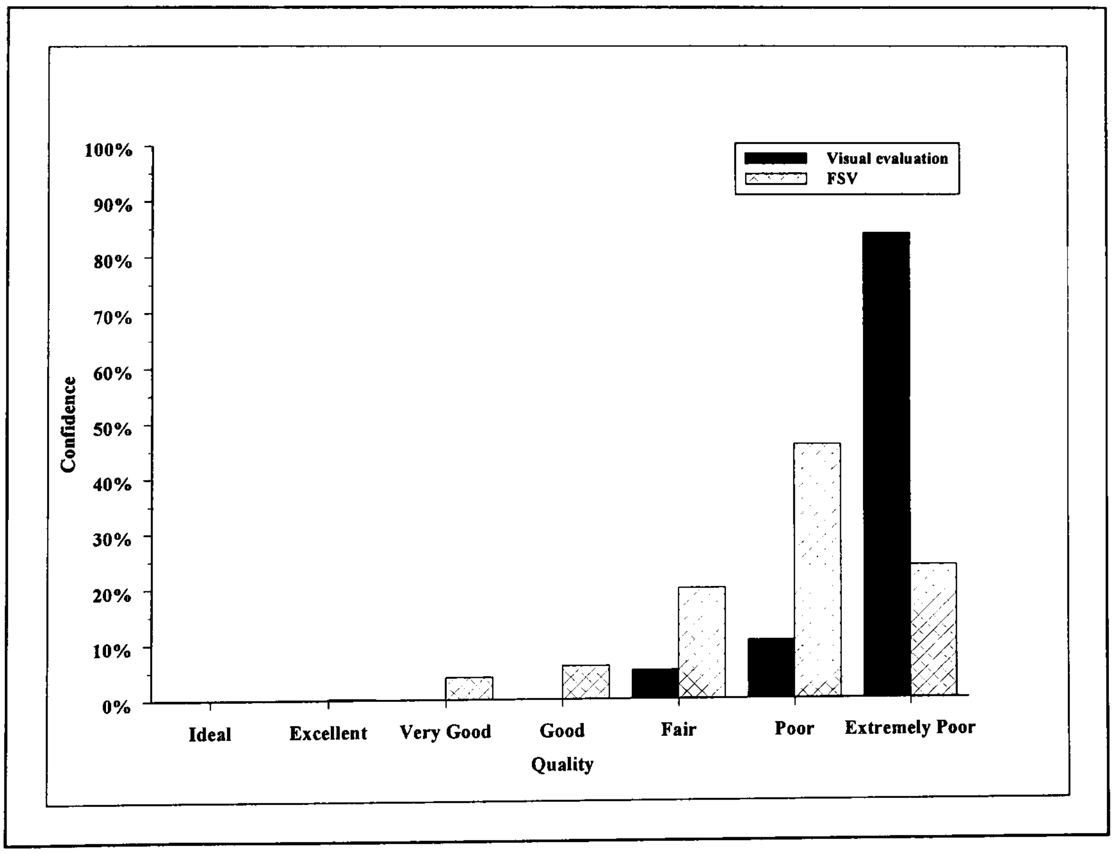

Figure 5.18: Visual / FSV confidence levels - comparison 4

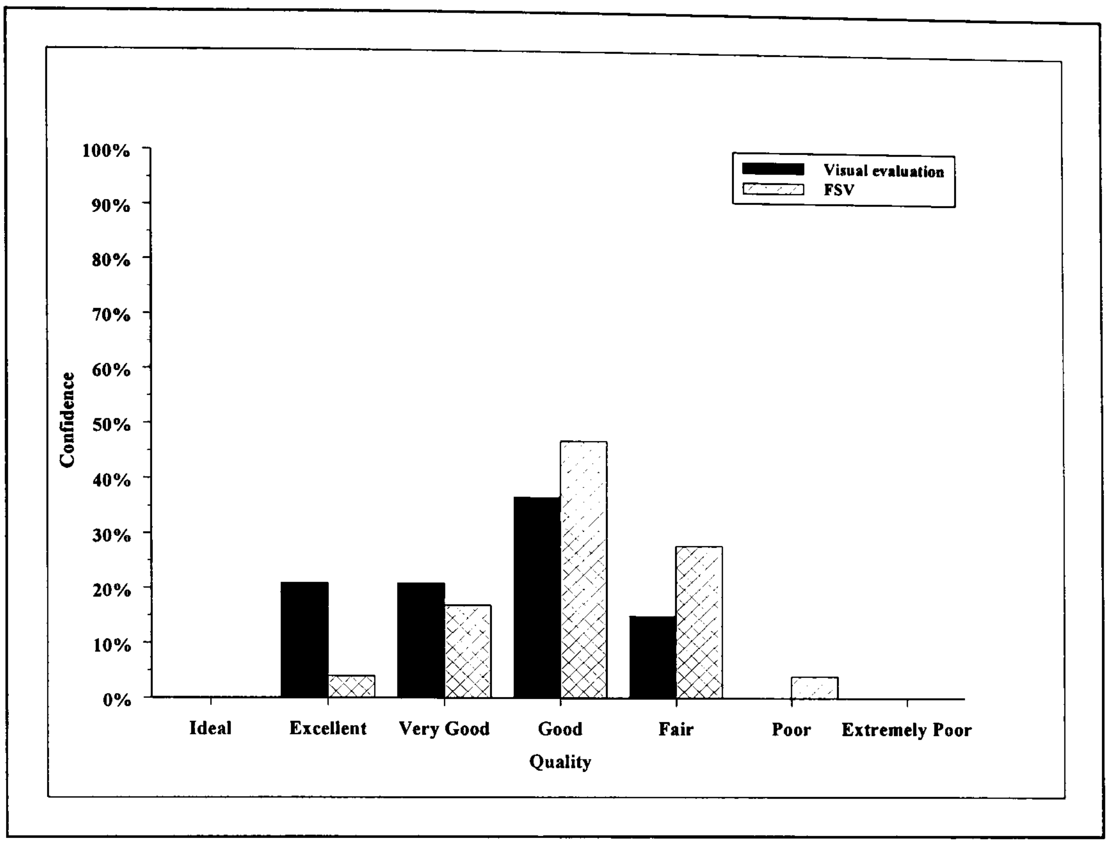

Figure 5.19: Visual / FSV confidence levels - comparison 5

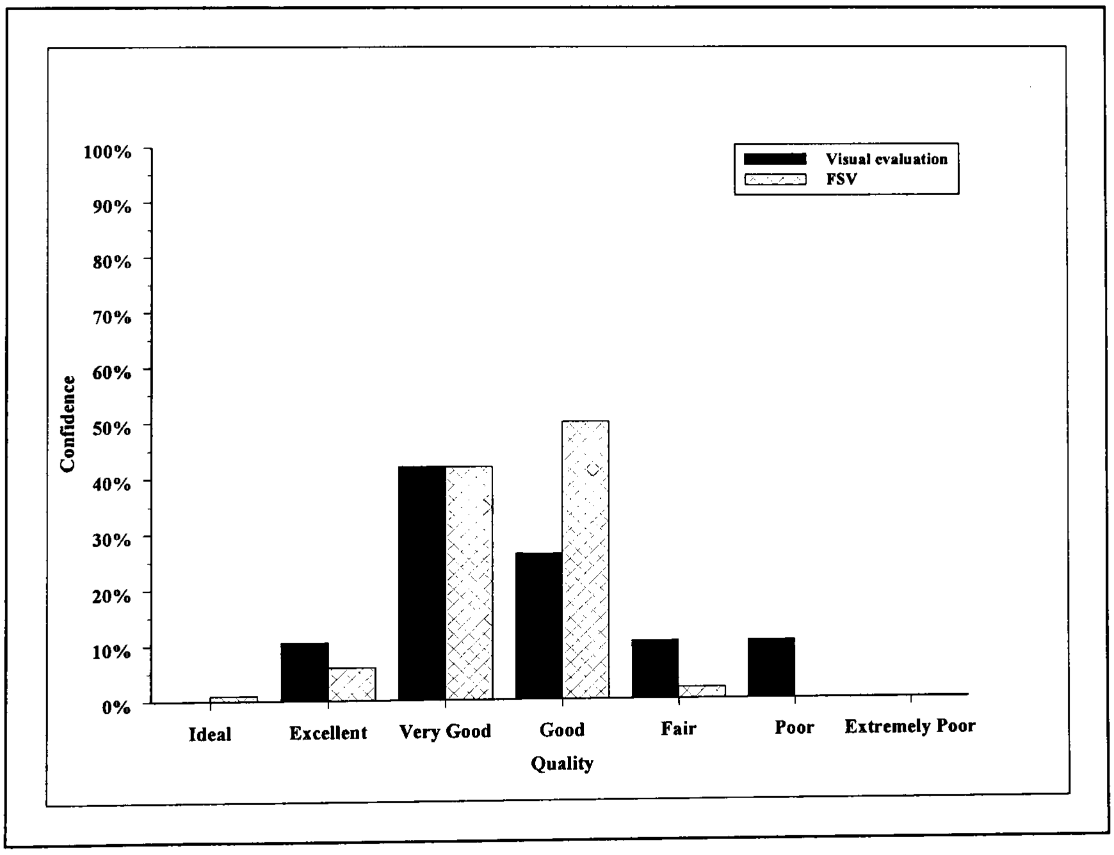

Figure 5.20: Visual / FSV confidence levels - comparison 6

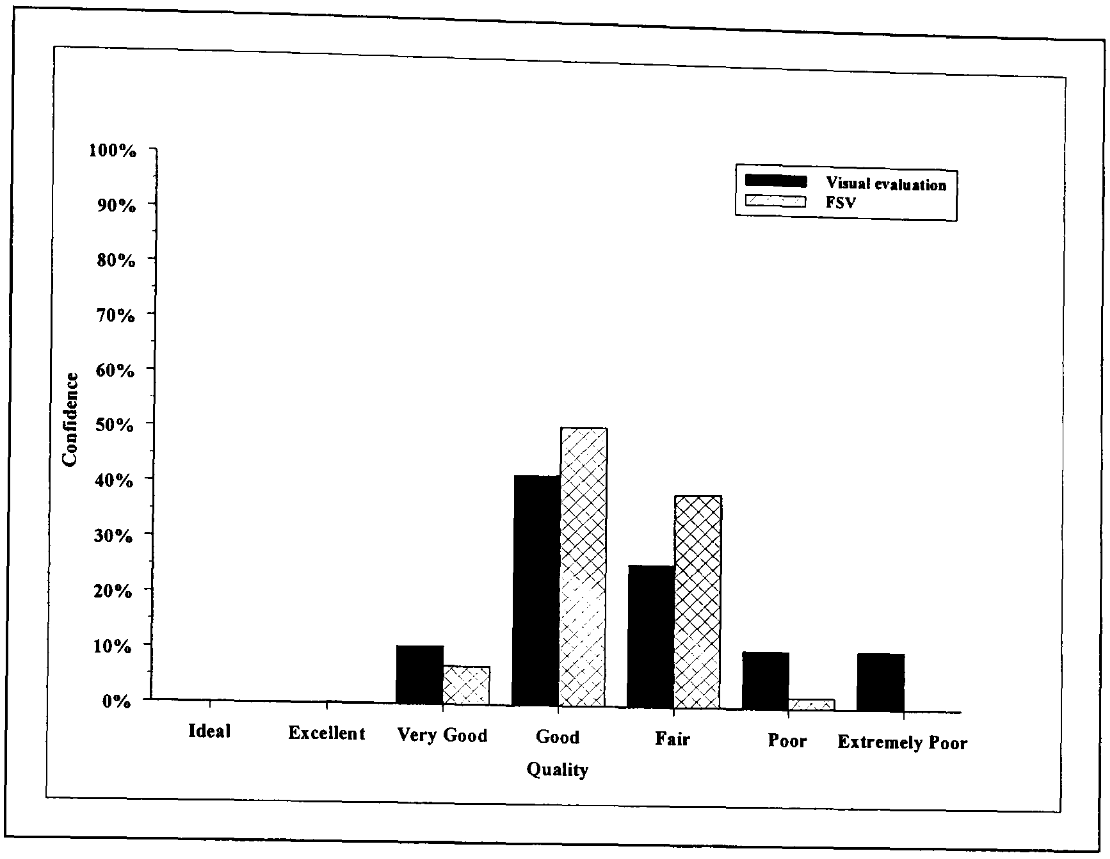

Figure 5.21: Visual / FSV confidence levels - comparison 7

5.5 CHAPTER SUMMARY

The results presented in this Chapter illustrate the Feature Selective Validation methods ability to mirror human perceptions in an assessment of compared signals. Validation information extracted employing the method is presented in a clear manner, employing a common scale used by engineers and scientists alike. The ability to weight the constituent measures which comprise an overall assessment of compared signals allows flexibility in situations where results from diverse areas of study are analysed. Whilst diagnostic information is readily available for in-depth investigations on the nature and magnitude of possible sources of error, allowing a rigorous and accurate analysis of the data acquisition method in question. However, in order to truly assess these potential advantages, the FSV method must be applied to several key areas of study. Chapter 6 details the application of the FSV method to three key areas of study, whilst reporting on the potential benefits of employing this quantitative validation method.

CHAPTER 6

FEATURE SELECTIVE VALIDATION CASE STUDIES

6. FEATURE SELECTIVE VALIDATION CASE-STUDIES

The Feature Selective Validation method possesses significant advantages over the automated validation methods of correlation, Zanazzi Jona, and van Hove (detailed in Chapter 5). In addition to the advantages held over current automated validation methods, the FSV method embraces the interpretation methods of visual evaluations employing a common quality scale, whilst producing confidence levels which mirror human interpretations of compared signal data sets. Further advantages are observed in the FSV method's ability to present high levels of diagnostic data based on the measures taken into account during a visual evaluation of results, whilst allowing a measured level of assessment flexibility in the form of subjective tolerances (weighting factors).

This Chapter aims to apply the FSV method to diverse application areas, illustrating the tremendous breadth of information obtainable through quantitative validation methods. The following Sections illustrate both the diagnostic capabilities of the FSV method and its ability to assist in the optimisation of methods employed to acquire complex data signals. In general, the acquisition of complex data signals may employ one of two methods: experimental (real); or modelled (computational or virtual). Analytical techniques are not covered in this Chapter, as, in general, these methods rapidly become intractable when employed to obtain complex data from highly complex application areas.

6.1 CASE-STUDY 1: EXPERIMENTAL REPEATABILITY

All experiments are subject to some inherent inaccuracy or loading and a detailed knowledge of experimental repeatability can assist in determining levels of acceptable experimental error. Validation of the results from one experimental technique against another, using the method detailed in this case-study, allows the potential benefit of assessing and more importantly locating common errors in both methods. This allows for the determination of experimental signature analysis, providing a rigorous framework for confidence building. Furthermore, validation techniques may be employed to identify best practice in a situation where there are ill defined test plans. This case-study illustrates how sets of experimental results can be compared and their differences quantified. Whilst, the interpretation of these results allows rational decisions to be made regarding the quality of measurement techniques or facilities, allowing them to be used with a measured level of confidence.

6.1.1 Theory

The quality of experimental data is influenced both by the method of producing and recording the data and the degree of perfection in the experimental procedure. For this reason, quantitative comparisons of experimental results are required to remove as much subjectivity as possible from the assessment of results. In electromagnetic measurements, such as EMC tests, complex signal data sets are common, making the development of validation techniques a complicated process. Tests repeated by different engineers, at different times or in different facilities will usually produce different results, further complicating the validation procedure. This invariably complicates the process of formally assessing whether differences between results are significant or acceptable. Clearly, a technique is required which removes the burden from the engineer of producing a quantitative analysis from a qualitative (usually visual) assessment. This case-study presents results from a number of repeated experiments and identifies their level of difference employing the Feature Selective Validation method detailed in Chapter 4. Further results investigate the origins of unreliability in the test procedures presented, classifying the inherent errors incurred and quantifying the magnitude of these errors.

6.1.2 Experimental Quality

It is conjectured that, acceptable repeatability may be set at a GDM value not greater than 0.2 (i.e. 'good'). The choice of this value is dependent on the inherent sensitivity of the measurements, and the application area under investigation. Within the FSV method, this value of GDM acceptability is defined as the Global Difference Tolerance (GDT), and is set to a value of 0.2 for the subsequent analyses.

6.1.3 Test Procedures

Three tests were repeated independently by two Engineers. All tests were based on the general instructions listed below for three different test structures:

1. ***Measurements were taken using a resonant cavity with a closely fitting, well defined, lid and an internal wire, used to excite the cavity. Terminations at both ends of the wire allowed the insertion loss of the cavity to be measured. The test plan involved the calibration of the cables and the fitting of the lid. This test should be highly repeatable as there are very few degrees of freedom.***

2. ***Resonance measurements were taken using a mode tuned cavity with internal 'paddle' tuner and a less well defined 'biscuit tin' lid. The test plan involved cable calibration, setting of the mode tuner and replacement of the lid. Here, the cavity is more poorly defined and there are more degrees of freedom with the test (lid fitting and paddle rotation).***

3. ***Radiation from the aperture resonator of Test 3 was measured using an external antenna. The test plan involved setting the aperture and external antenna approximately 1m apart with the aperture approximately 50cm below the centre line of the antenna and angled at 30° to the vertical and the antenna rotated by 15° from the vertical (no guidance was issued as to whether the rotation should be clockwise or anticlockwise). This was an ill defined test with a large number of degrees of freedom, principally the rotation of the receiving antenna. The test was carried out in a general laboratory and no guidance was issued on the location of the test.***

Essentially the tests ranged from a well defined to a poorly defined system.

6.1.4 Initial Results

Results from the test procedures described previously are illustrated in Figures 6.1 - 6.3.

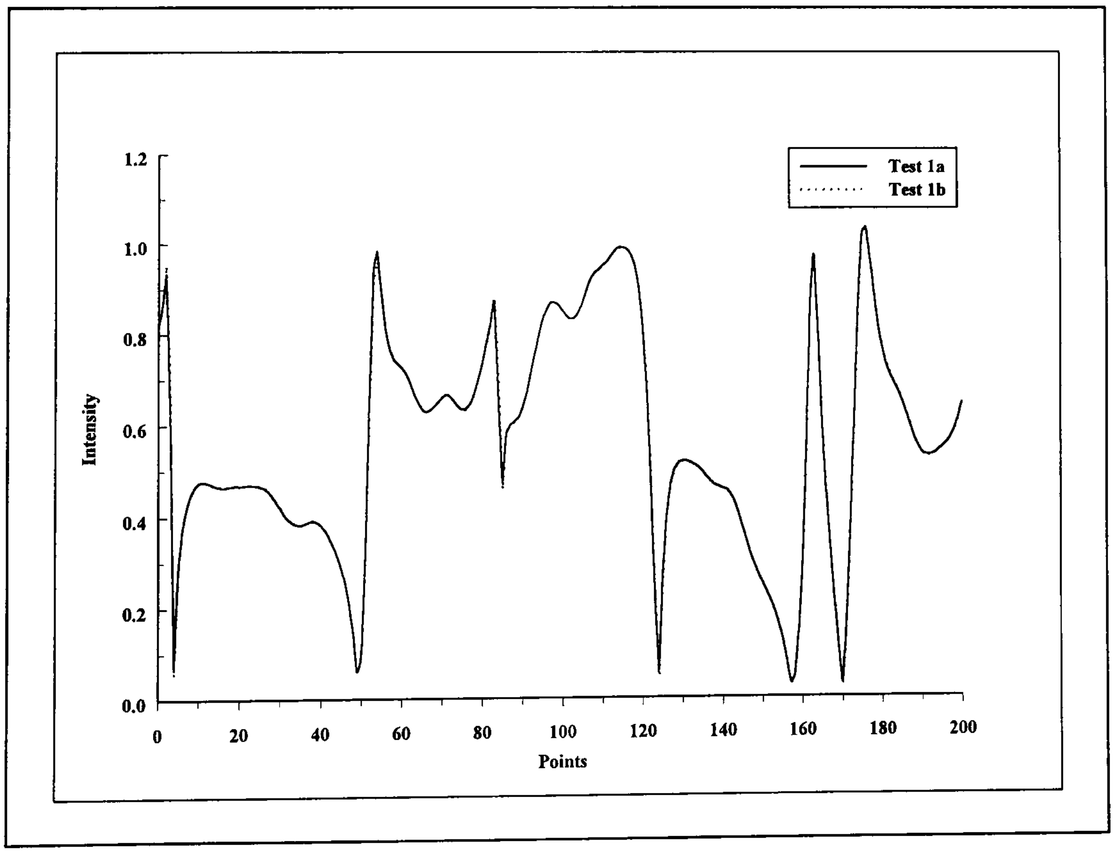

Figure 6.1: Comparison of repeated results - Test 1

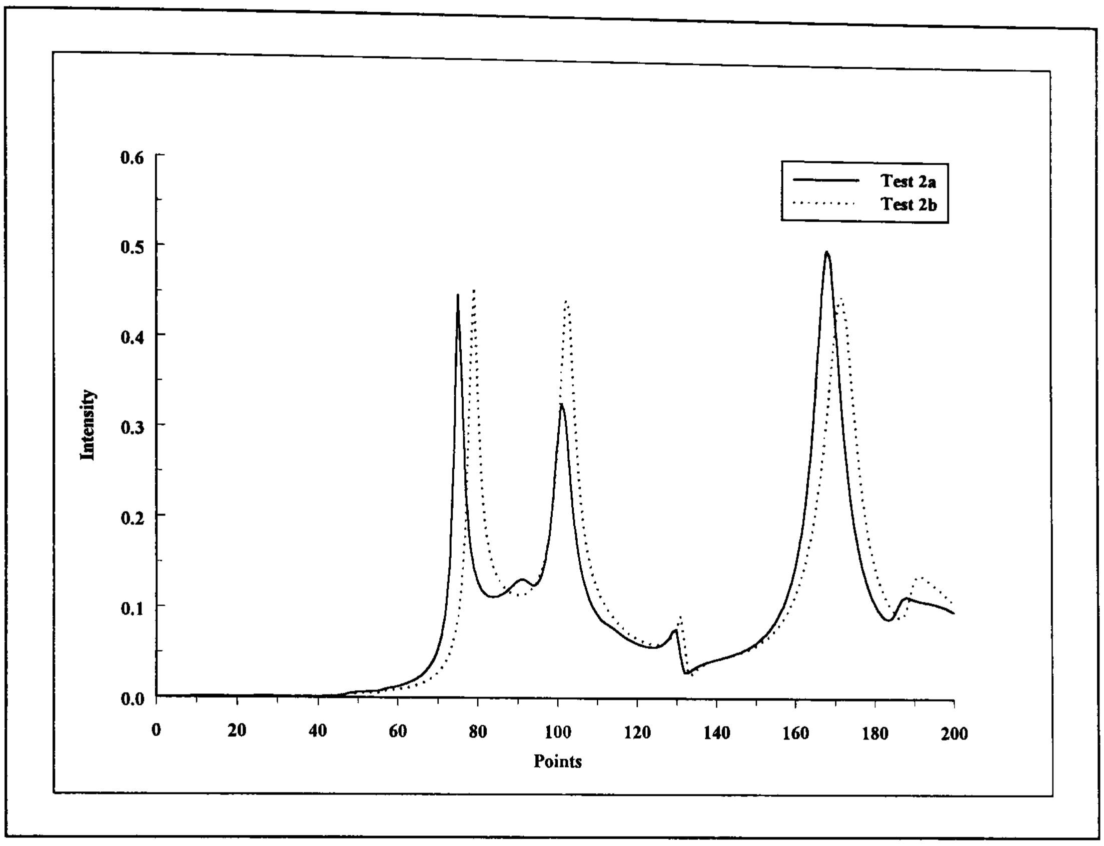

Figure 6.2: Comparison of repeated results - Test 2

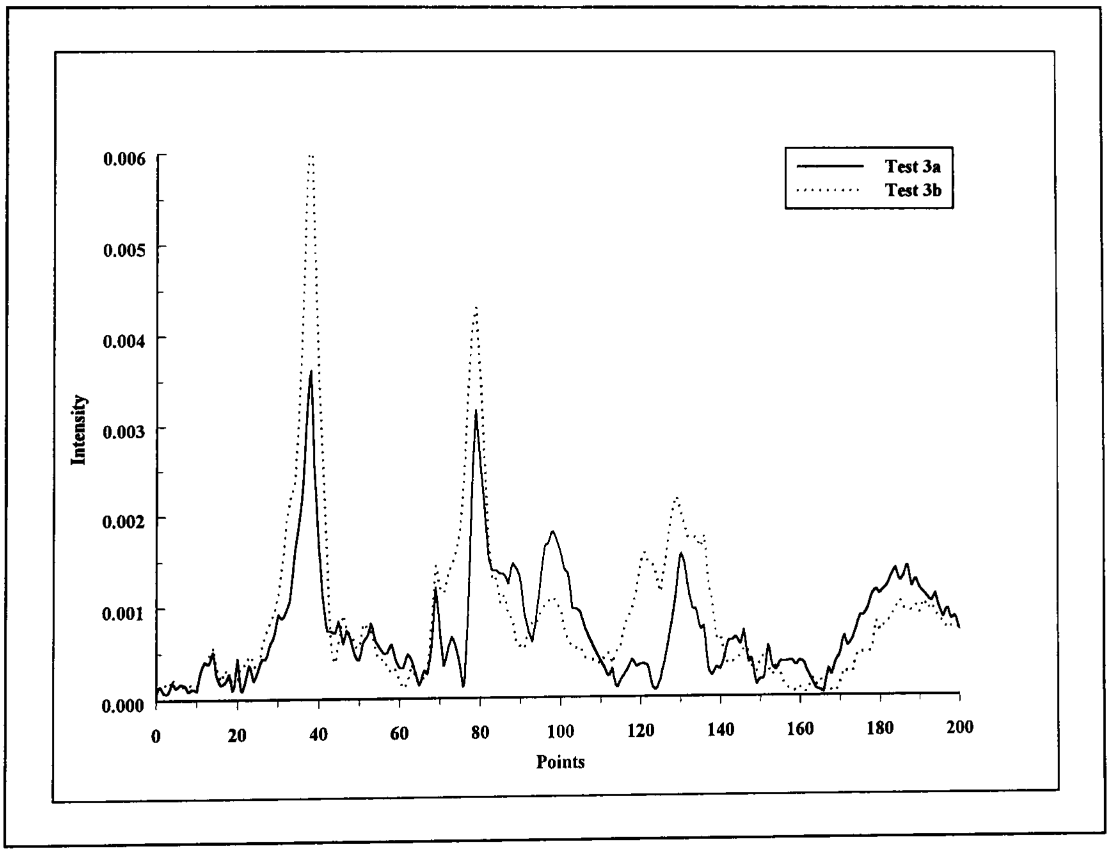

Figure 6.3: Comparison of repeated results - Test 3

All tests were performed as indicated in Section 6.1.3 and the GDM was evaluated for the three test cases. The results are summarised in Table 6.1.

Test	***ADM***	***FDM***	***GDM***	***Assessment***
Test 1	*0.01*	*0.01*	*0.01*	*Ideal*
Test 2	*0.06*	*0.23*	*0.24*	*Good*
Test 3	*0.24*	*0.25*	*0.39*	*Fair*

Table 6.1: Summary of results - Tests 1 to 3.

Initial results indicated in Table 6.1 and the respective comparisons of Figures 6.1 - 6.3 illustrate that Test 1 achieves a high standard of reliability with very little discrepancy between the two sets of results, indicated by a GDM value of 0.01 or 'ideal'. Results obtained employing Test 2 indicate a clear difference between the two sets of results, although the GDM value of 0.24 is not sufficient to imply that the results are from different structures. The results obtained employing Test 3 however, indicate major discrepancies between the compared results, with the value of the GDM indicating that the results may be from the same type of test procedure, but not necessarily the same test structure.

6.1.5 In-depth Analysis

Figures 6.4 - 6.6 illustrate the corrected results of test procedures 1 - 3 employing the FSC method. Furthermore, the correction of distortions within the results allows valid quantification of the magnitude and class of error acting on the results illustrated in Figures 6.1 - 6.3. The results of this study are indicated in Table 6.2, comprising structural, positional and global repeatability.

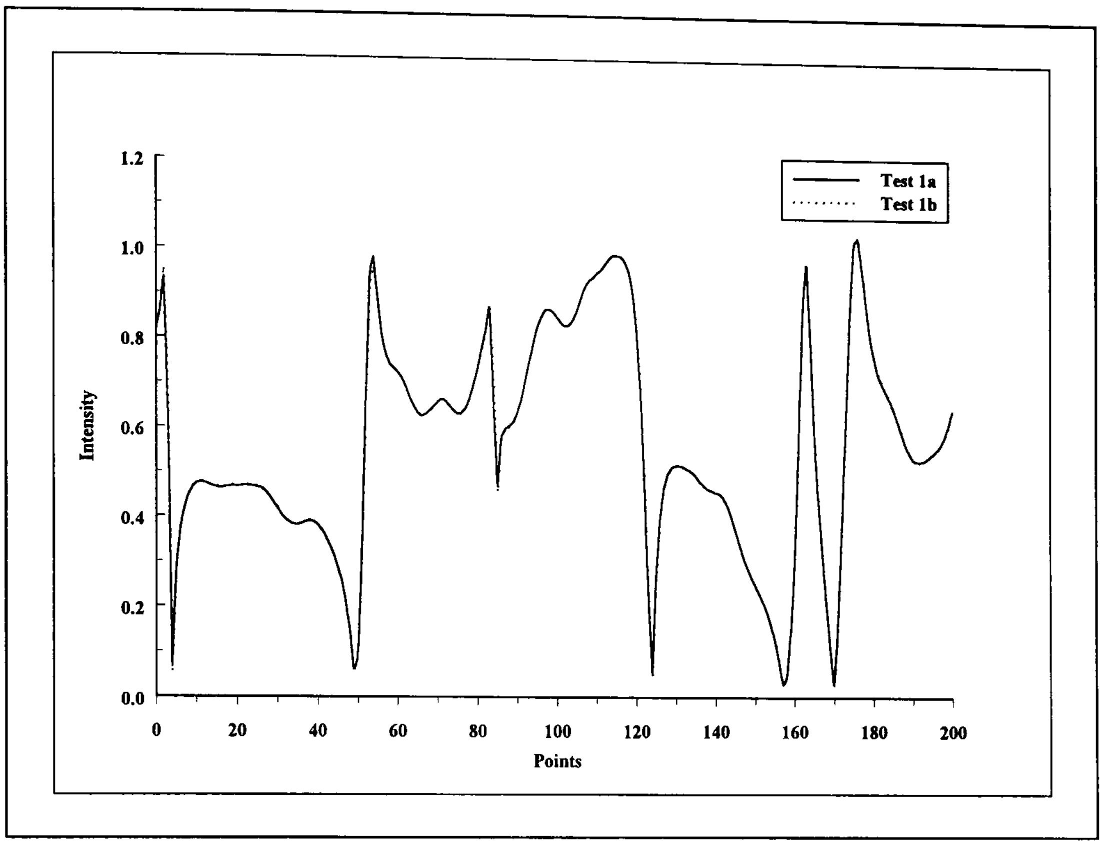

Figure 6.4: Corrected results - Test 1

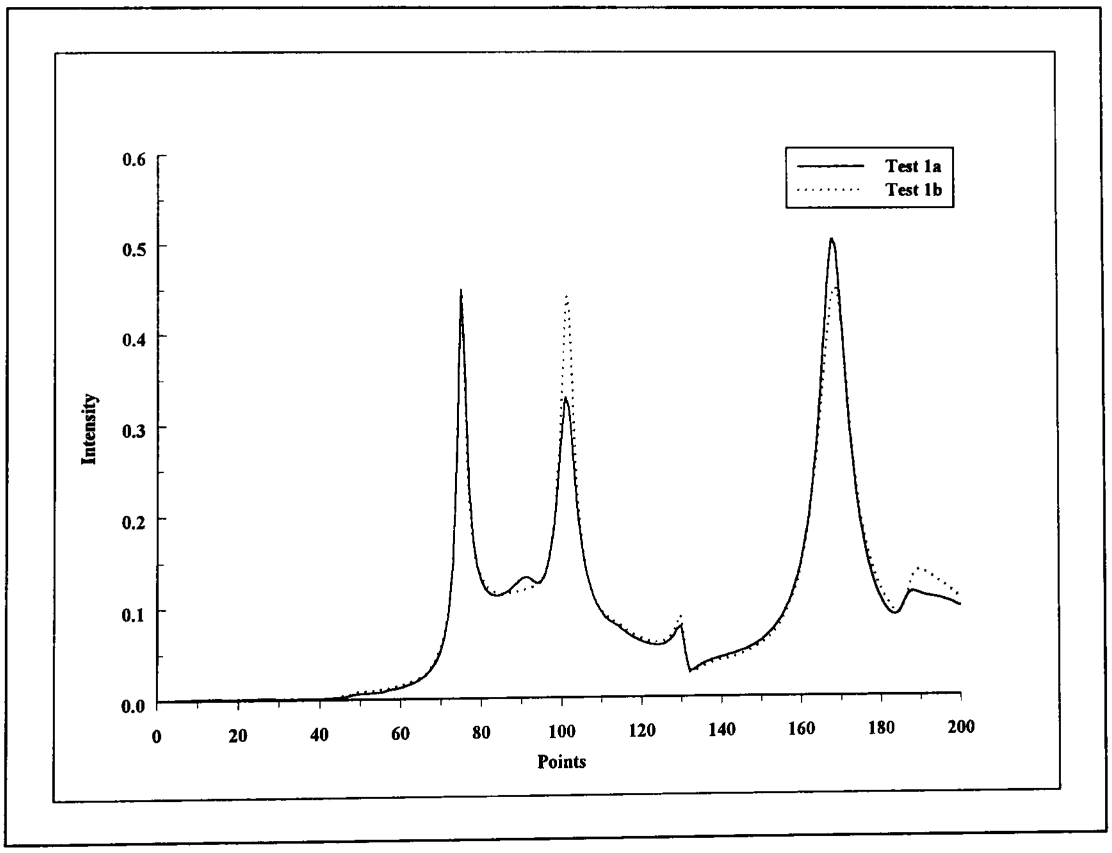

Figure 6.5: Corrected results - Test 2

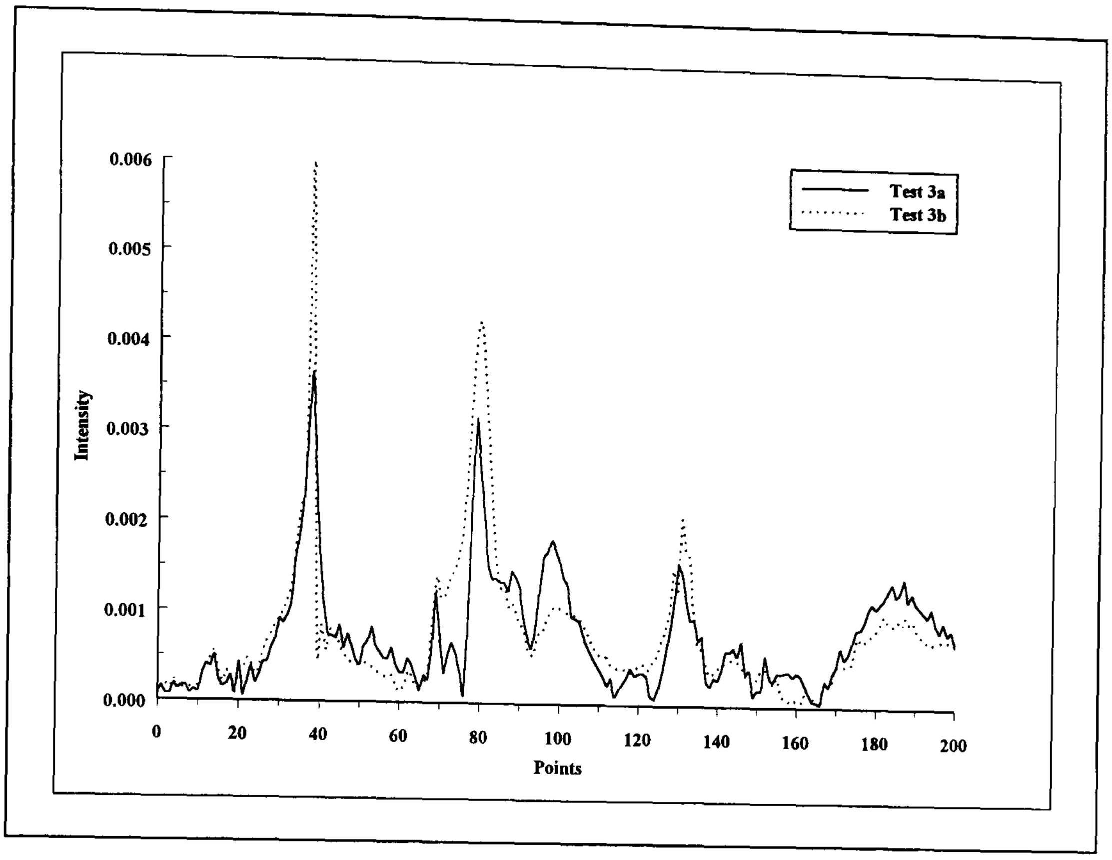

Figure 6.6: Corrected results - Test 3

The global repeatability values of Table 6.2 are taken from Table 6.1, and indicate the overall or uncorrected ability of each test structure to produce repeatable results. However, for diagnostic purposes this measure may be separated into two constituent parts namely; structural and positional. That is to say that after the compared signals have been corrected, a new or undistorted measure of repeatability is obtained, with the difference between uncorrected and corrected repeatability indicating the distortion repeatability within the comparison data sets. Furthermore the structural measure of repeatability (corrected GDM) can be subdivided into three components namely; amplitude structure, feature structure and global structure. The values illustrated in Table 6.2 indicating amplitude, feature and global structure differences denote the repeatability of the: general amplitude levels; feature shapes; and overall feature shapes and amplitude levels respectively within the three test cases of Section 6.1.3. Furthermore, these measures are independent of any incurred distortion effects between signals and are obtained from an analysis employing the aligned or corrected comparisons of Figures 6.4 - 6.6.

	Discrepancies				
	Structure			*Position*	*Global*
Test	*Amplitude*	*Features*	*Global*	α	*GDM*
Test 1	*0.01*	*0.01*	*0.01*	*0.00*	*0.01*
Test 2	*0.02*	*0.06*	*0.07*	*0.17*	*0.24*
Test 3	*0.15*	*0.20*	*0.26*	*0.13*	*0.39*

Table 6.2: Classification of structural, positional and global repeatability

6.1.6 Results

Employing the FSV method, accurate quantitative/qualitative assessments based on the repeatability of test structures and procedures may be made. It is noted from the results, as expected, that well defined experimental structures and procedures achieve greater repeatability in comparison to those of a less well defined nature. However, all results can be quantified and a value expressing the level of agreement or disagreement can be quoted with confidence. Using the suggested level of acceptability (less than 0.2) and the uncorrected results of Figures 6.1 - 6.3, Test 1 would be regarded as acceptable, Test 2 may be regarded as unacceptable (but a borderline case in need of further investigation) and Test 3 is clearly unacceptable.

However, employing an in-depth analysis of uncorrected and corrected data, repeatability within each of the test procedures described in Section 6.1.3 can be subdivided into isolated areas of discrepancy. Using the suggested level of acceptability of 0.2, test procedures 1 and 2 exhibit acceptable levels of repeatability for both structural and positional (distortion) characteristics. However, test procedure 3 whilst exhibiting a repeatable value of positional error is unacceptable (but a borderline case in need of further investigation) in terms of structural error.

Using a similar approach, the quality of different test facilities, the approach of different technical staff and of test instructions or plans can be assessed. From these, a rigorous framework for identifying and disseminating good working practice may be constructed.

6.2 CASE-STUDY 2: MODEL OPTIMISATION

Although it is tempting to continually refine the structure of numerical models in order to produce increasingly accurate results, there comes a point at which further developments produce imperceptible changes. A typical example is the reduction in mesh size for a simulation employing the TLM method[Christopoulos 1995] such that it becomes a better approximation to a physical structure. One way of estimating the maximum allowable node compliment is to determine the maximum acceptable run time and calculate back to evaluate the maximum number of nodes and, therefore, the required resolution. However this approach may be invalid because the final resolution may be greater than necessary, in which case the computing facilities will be inappropriately used; or the resolution may not be adequate, in this case the results may not be reliable. Employing the FSV method however, results from several model resolutions may be compared, allowing the assessment of optimum discretisation levels for individual models based on the convergence of the validation results gained. This case-study illustrates how sets of modelled results can be compared and their differences quantified in a search for the most efficient resolution. The method described is based on a single figure of merit (GDM) employed to identify the level of difference obtained from a comparison of two or more sets of results. Comparisons may now be made between results on a quantitative, rather than qualitative basis. This case-study shows how the interpretation of these results allows rational decisions to be made in the trade-off between run time, memory requirements, resolution and quality.

6.2.1 Introduction

Experimental procedures consist of flexible arrangements of physical components with precise characteristics and geometries known to an engineer. Conversely, numerical techniques or simulations use models constructed from available building blocks based on numerical relationships or equations. These model the behaviour or characteristics inherent in specific experimental design based problems and allow predictions of performance to be made. By reassessing the model equations or attributes of the system, it is perceived that the required performance can be obtained. In order to be useful, numerical models invariably involve simplifications: assumptions concerning the characteristics of the simulated model are made; small effects are neglected; and idealised relationships are assumed.

A numerical techniques primary objective is to simulate or reproduce the behaviour of certain stimuli under specific conditions constrained by space and time. This is accomplished by a set of individual elements (or building blocks) that may be linked to form models of real structures (experiments). Invariably limitations are placed on each of the individual elements and the fully formed model by the computational and memory inadequacies of platforms on which simulations are run. This case-study investigates the application of the FSV algorithm in identifying the optimum construction of a simulation or model. The FSV method compares two signals and produces information expressing their differences. The method is illustrated by the mesh size refinement of a simple TLM method based resonator model. It shows how the GDM changes with varying mesh refinements and identifies the optimum mesh size based on a consideration of run time, memory requirement and the accuracy of the model. As a consequence, a level of accuracy can be attributed to simulations of a more granular structure. Although the discussion is specifically related to TLM, the method could be translated to other numerical modelling methods.

6.2.2 Theory

In order to develop a quantitative method for assessing model refinements within a simulation package, it is imperative that an understanding of how the general limitations of computational platforms tend to degrade the performance of simulation packages. Memory allocation sets the limits of a simulation in terms of both the space and time steps on which a model is evaluated. This inherent problem is known as discretisation. Hence, if a model is simulated employing the infinitely small steps in space and time (δl and δt) of the real world, the platform on which it runs would require an infinite number of memory locations for the spatial calculations and an infinite number of time steps for the transient calculations.

Obviously this is not practically attainable, nor is it desirable, as the main reason for developing modelling techniques is to simulate a real structure as closely as possible, without the complications, cost and unwieldiness of the real thing. To resolve the problem of limited memory and time, the infinite steps in space and time are approximated by the use of a grid or mesh to Δl *and* Δt respectively. The extent to which these limitations tend to degrade the capabilities of a numerical evaluation of a real structure are better understood through the scrutiny of a single modelling technique. A typical example is the reduction in mesh size of a numerical TLM method model so that the model becomes a better approximation to the physical structure. However, a decrease in node size by a factor of two increases the number of nodes required to construct a 3D model by a factor of eight (resulting in an increase in memory requirement and run time).

6.2.3 Test Structure

The structure used to asses the performance of the method detailed in this case-study is illustrated in Figure 6.7, with the nearest corner removed for clarity. It shows a resonant cavity of dimensions 200 mm (x) x 300 mm (y) x 180 mm (z). A wire of diameter 10 mm runs along the length of the cavity and is used to excite the structure and simultaneously measure the signal modified by the presence of the cavity.

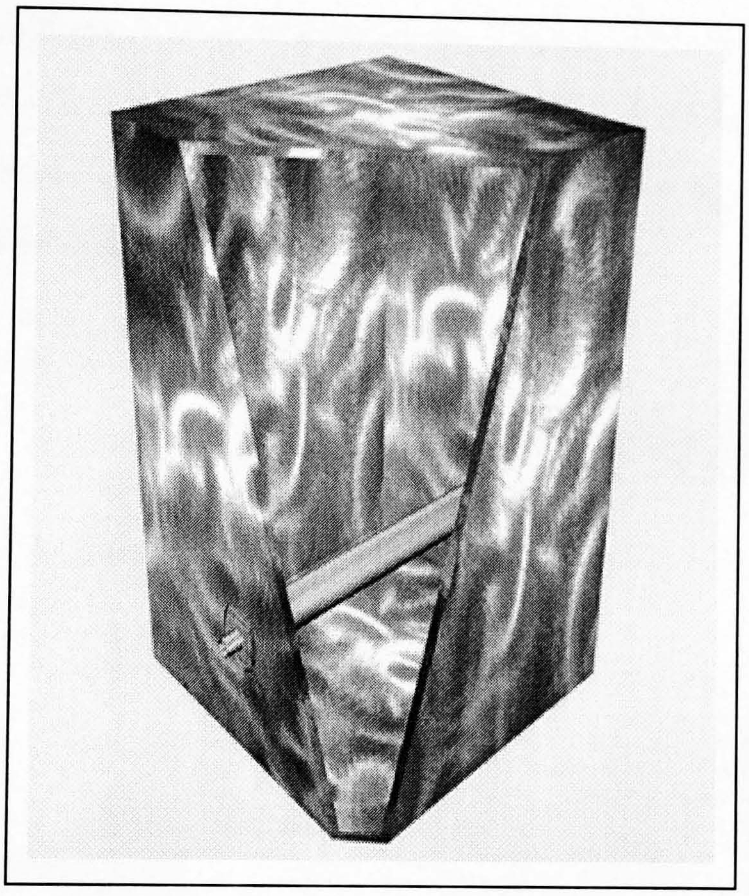

Figure 6.7: Test structure

6.2.4 Method

The illustrative method detailed in this case-study employs the FSV algorithm, and more specifically, the GDM value obtained from a comparison of complex signals, along with both the node compliment and node size of the numerical model under investigation. Furthermore, the graphical representation of the three parameters detailed previously along with a suggested maximum allowable simulation node compliment and Global Difference Tolerance (GDT) allows for an evaluation of the optimum discretisation area of a numerical model as a function of numerical simulation accuracy and the computational constraints of the platform used. This effectively reduces the subjectivity normally associated with this area of study.

6.2.5 Results

Simulation results were obtained using the test structure of Figure 6.7, employing hybrid symmetrical condensed nodes of 10 mm, 20 mm, and 30 mm. Results obtained using 10 mm symmetrical condensed nodes were below the nominal $\lambda/10$ discretisation level and were used as a reference for all comparisons. Figure 6.8 illustrates the simulation results obtained, whilst Table 6.3 indicates the comparisons along with there respective node compliments and GDM values.

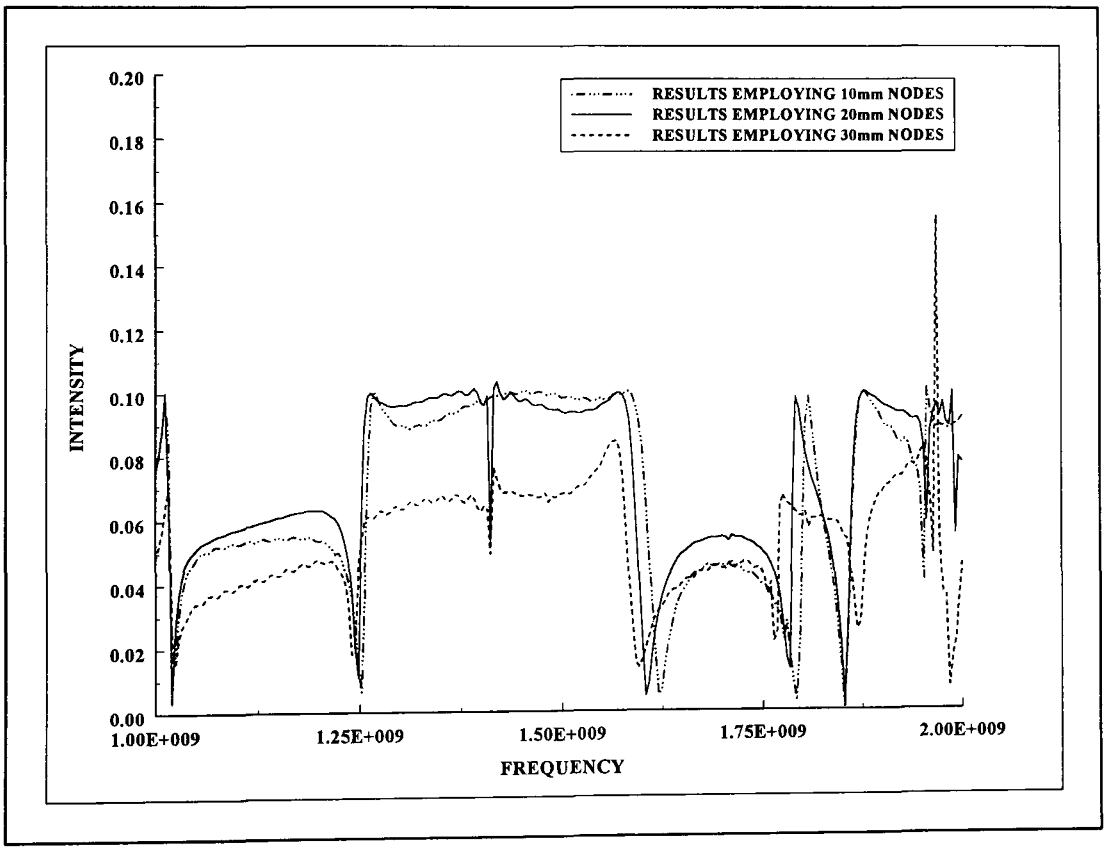

Figure 6.8: Results employing 10 mm, 20 mm and 30 mm nodes

Comparison	*Node Compliment*	*GDM*
10 mm / 10 mm	***10,800 - (10 mm)***	***0.00***
10 mm / 20 mm	***1,760 - (20 mm)***	***0.29***
10 mm / 30 mm	***616 - (30 mm)***	***0.54***

Table 6.3: Node compliment and GDM values for comparisons

A graphical representation of the results indicated in Table 6.3 is illustrated in Figure 6.9, where the primary Y-axis represents the node compliment of the numerical simulations. The secondary Y-axis represents the Global Difference between simulation results equated using the FSV method, whilst the abscissa represents the node sizes of the numerical models under investigation. Furthermore, the left and right horizontal dotted lines represent the suggested maximum allowable node compliment and GDT respectively. When several sets of comparisons are evaluated and graphically displayed against their respective node sizes and node compliments, a clear indication of the maximum allowable discretisation levels can be obtained as a function of computational resources and numerical simulation accuracy.

The area of the graphical representation encompassed by the solid and dotted traces shown in Figure 6.9 indicates the discretisation levels achieved using the computational platform in question, along with their respective node compliments and accuracy. Whilst the area of the graphical representation encompassed by the vertical dotted lines represents the suggested optimum discretisation levels for the model under evaluation. The optimum discretisation area is shown in Figure 6.9 as a solid arrow. Furthermore, using a suggested maximum node compliment of 4000 nodes and GDT of 0.4 or 'fair', the optimum discretisation area achieved is 17.5 mm to 24.25 mm.

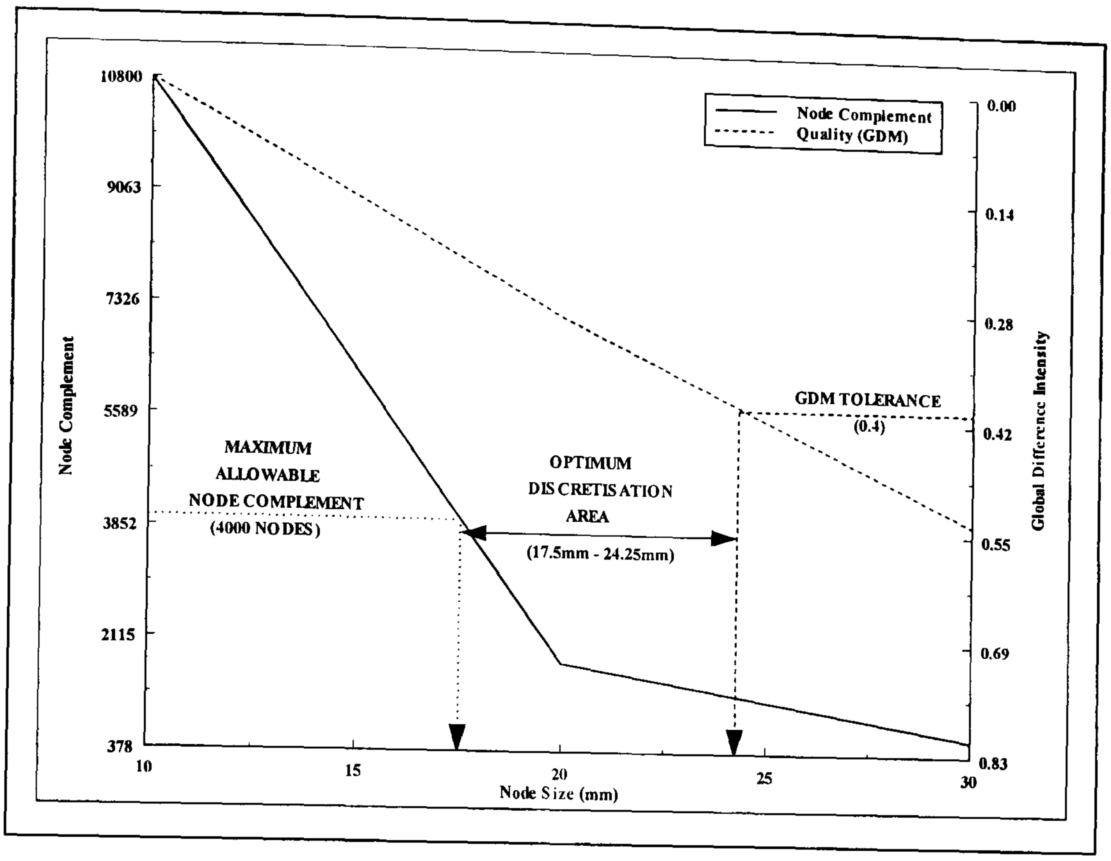

Figure 6.9: Model optimisation plot

If a specific numerical modelling resolution is chosen, both the quality of the results gained from the simulation and the node compliment of the respective model may be approximated, allowing a rigorous and focused investigation into a single, optimum model resolution. Hence for a node size of 18 mm, the node compliment is equated to 3600 nodes and the approximated accuracy of the simulation is 0.24 or 'good'.

6.2.6 Discussion

Optimisation results have been presented employing a graphical method which incorporates the versatility of the FSV method and more specifically the GDM of several comparisons, along with a predefined GDT. This technique allows the optimum discretisation level of individual numerical models to be quantitatively assessed. It is noted from the results, as expected, that numerical models employing course meshes exhibit less reliable results than those employing fine meshes, although there does come a point at which a reduction in node size may not be necessary for particular applications.

The implementation of the graphical procedure illustrated in this case-study employing the FSV method allows for the optimum discretisation area of the model under investigation to be quantitatively deduced, allowing computational facilities to be used efficiently. Using the method described in this case-study, a rigorous framework may be constructed for identifying and disseminating good working practice among engineers employing numerical modelling techniques.

6.3 CASE-STUDY 3: DNA CORRECTION AND IDENTIFICATION

An almost unparalleled degree of certainty in the uniqueness of biological samples collected from different locations at different times may be obtained from the analysis of ***Deoxyribonucleic Acid*** (***DNA***) fingerprints. Theoretically DNA technology offers to crime investigators and the courts an opportunity to characterise "without reasonable doubt" an individuals biological 'serial number', providing solid proof of an individuals 'innocence'. However, many factors affect both the DNA samples collected from crime scenes and the information obtained from the analysis of these samples.

This case study aims to apply the Feature Selective Validation (FSV) method and more specifically the Global Difference Measure (GDM) to several DNA sequences in an attempt to identify and group sets of biological samples. The Feature Selective Correction (FSC) method is employed to correct distortions between DNA sequence data sets before judgement on the quality of a comparison is made. The FSC method is applied to remove distortions caused by the employment of different gel types used to extract DNA sequences from biological samples. The results present a clearly defined analysis technique which may be employed to justifiably remove unwanted characteristics inherent in DNA sequence samples, whilst allowing a true evaluation of the fitness of one DNA fingerprint to another.

6.3.1 Theory

If presented correctly, a high level of confidence is associated with the information obtained from DNA fingerprint analysis. However, correct or accurate presentation of DNA data is inhibited both by the method of biological sample collection and the method by which DNA sequences are extracted from these samples. The integrity of DNA data may be degraded by the method employed to collect initial biological samples, the method employed to store these samples[Balazs 1990], and the technique employed to extract DNA sequences from collected biological samples. The simplest method of DNA extraction employs blood as the biological sample source, however the best source of cellular material is obtained from bone marrow. The optimum temperature at which biological material should be stored is -70°C or over liquid nitrogen. Further to the sampling and storage requirements of biological materials, methods employed to extract DNA sequences from biological samples will invariably distort the characteristics of a DNA fingerprint. A multitude of chemicals (e.g. sucrose, magnesium chloride, phosphate buffered saline (PBS), sodium lauryl sulphate) and machinery (e.g. centrifuge, blender, mortar) are employed in the process of releasing DNA from biological samples. Finally, the inherent characteristics of a DNA sample must be realised graphically, and again multiple techniques are available to obtain this data.

Clearly the method of collecting and storing biological samples and extracting DNA sequences from these samples is a complicated process and variabilities will inevitably occur. These variabilities however, will not only occur between DNA fingerprints of different individuals, but also those processed from the same subject. Furthermore, this problem is exacerbated in circumstances where the reference and comparison DNA sequences are extracted at different times employing different techniques. An example of this problem is the comparison between a DNA fingerprint extracted from biological matter collected from a suspect and a DNA fingerprint extracted five to six years previously from biological matter collected from a crime scene. In this example, a straightforward comparison of the two DNA sequences without distortion correction would rely on the techniques employed at the extraction stage of these DNA sequences being exactly the same.

6.3.2 DNA Data

Historically, DNA data has been presented as profiles or bands of different grey scale intensities, however, from these profiles data of a one dimensional nature (line graph) may be obtained. It is from this one dimensional data that quantitative investigations as to the identity of DNA structures may be made. The data illustrated in Figures 6.10 and 6.11 indicate the inherent DNA characteristics of two birds. Figure 6.10 represents DNA data extracted using two different gel types for a male bird, whilst Figure 6.11 represents DNA data extracted from a female bird, presented on two different gel types.

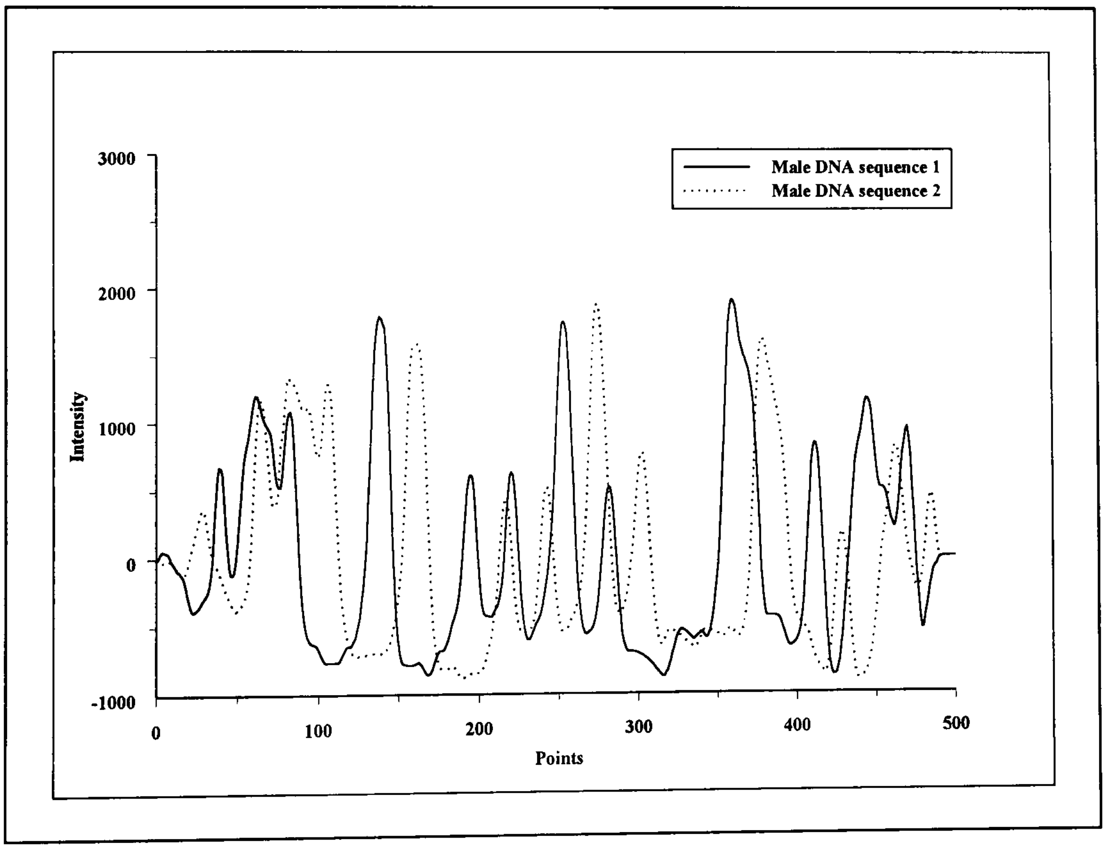

Figure 6.10: Comparison of uncorrected male bird DNA sequences

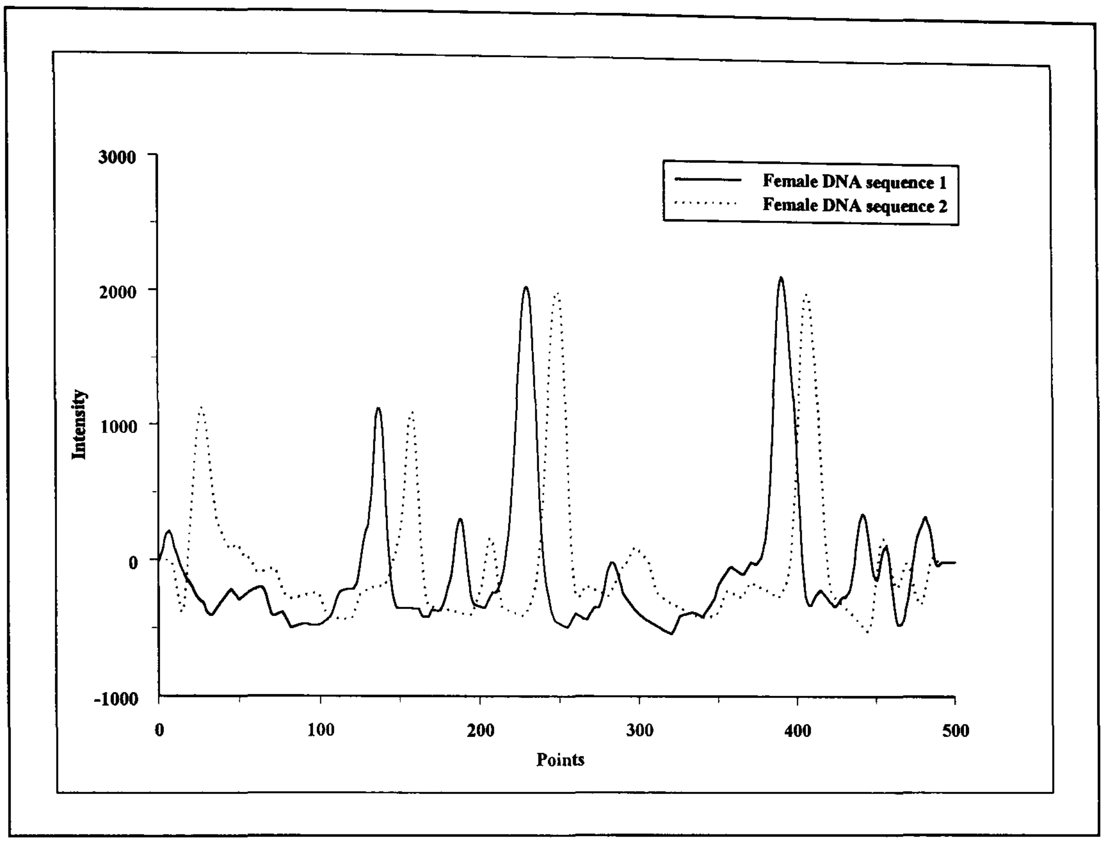

Figure 6.11: Comparison of uncorrected female bird DNA sequences

An initial visual inspection of Figures 6.10 and 6.11 indicates severe shifts and stretches between the pairs of DNA data. These distortions between potentially identical sets of data are due to different gel types employed to extract the DNA fingerprints from the biological samples provided. Although the main distortion between data sets is a single linear shift of approximately 25 - 30 points, more complex shifts and stretches may be embedded in the data sets. Figures 6.12 and 6.13 illustrate corrected versions of the comparisons illustrated Figures 6.10 and 6.11 obtained employing the FSC method. Significant improvement over the original comparisons are observed, with the elimination of the 25 - 30 point shift inherent in the results of Figures 6.10 and 6.11.

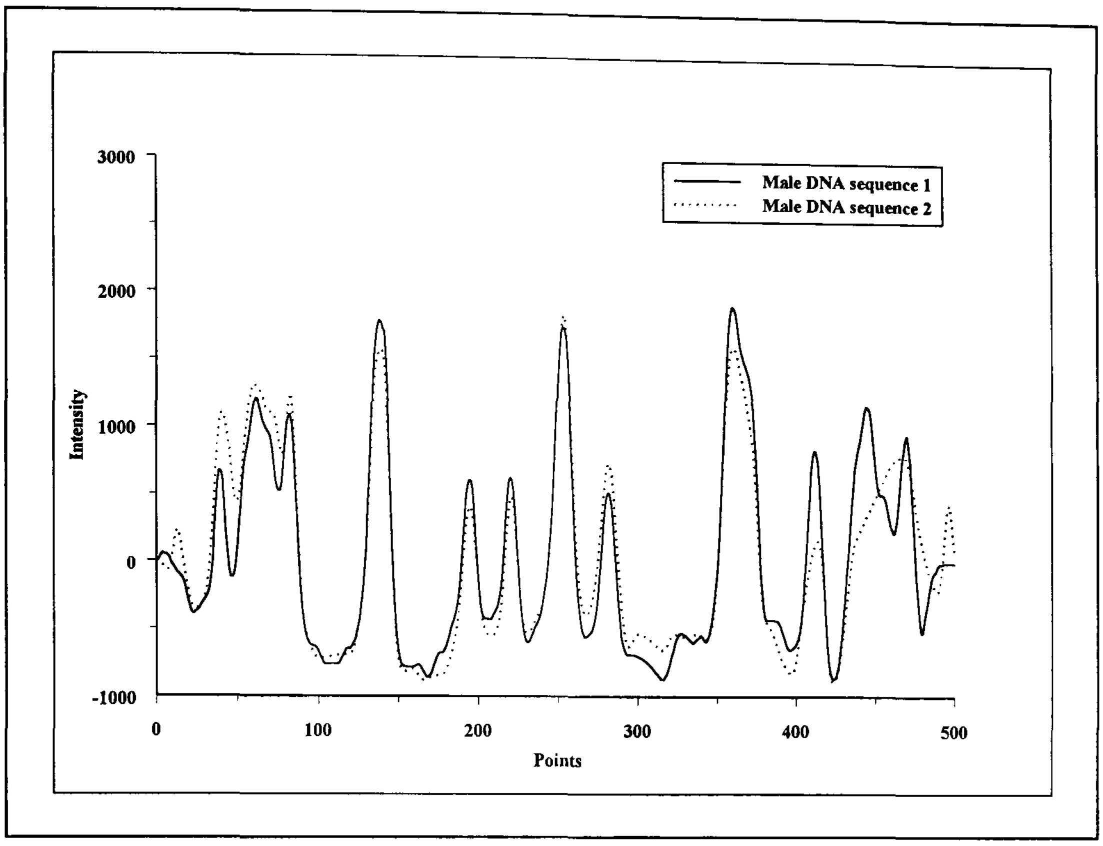

Figure 6.12: Comparison of corrected male bird DNA sequences

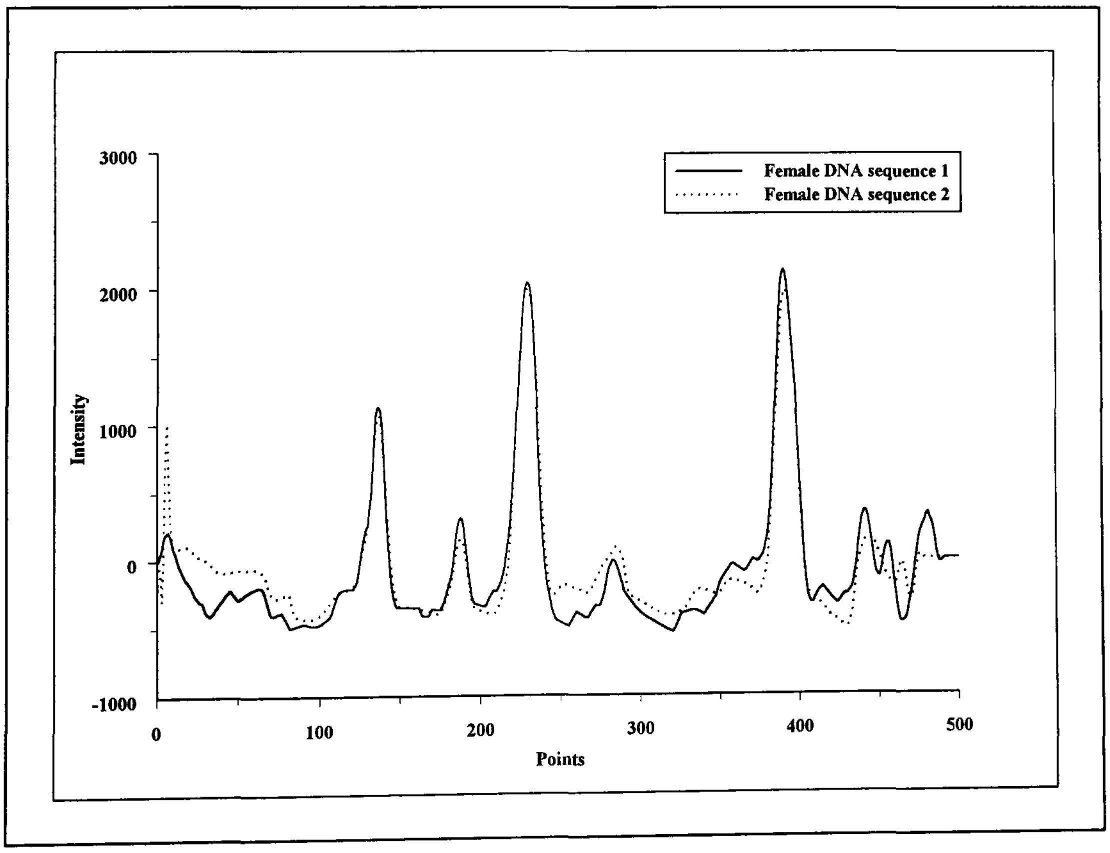

Figure 6.13: Comparison of corrected female bird DNA sequences

6.3.3 Results

Tables 6.4 and 6.5 indicate the FSV results obtained from an analysis of the uncorrected and corrected DNA results illustrated in Figures: 6.10 and 6.11; and 6.12 and 6.13 respectively.

	ADM	*FDM*	*GDM*
Uncorrected	*0.39*	*0.56*	*0.78*
Corrected	*0.09*	*0.11*	*0.15*

Table 6.4: Male/Male bird DNA comparison (sequence 1 - sequence 2)

	ADM	*FDM*	*GDM*
Uncorrected	*0.34*	*0.50*	*0.65*
Corrected	*0.09*	*0.12*	*0.18*

Table 6.5: Female/Female bird DNA comparison (sequence 1 - sequence 2)

The results of Tables 6.4 and 6.5 indicate clear improvement between the uncorrected and corrected DNA comparisons of potentially identical birds. With the FSV results indicating initially 'poor' comparisons for both the male and female DNA sequences. However, employment of the FSC method increases the confidence associated with the male and female DNA sequences indicating 'good' comparisons.

Tables 6.6 and 6.7 indicate the uncorrected and corrected cross DNA sequence comparisons of male and female birds.

	ADM	*FDM*	*GDM*
Uncorrected	*0.35*	*0.54*	*0.76*
Corrected	*0.22*	*0.32*	*0.46*

Table 6.6: Male/Female bird DNA comparison (sequence 1)

	ADM	*FDM*	*GDM*
Uncorrected	***0.37***	***0.51***	***0.78***
Corrected	***0.23***	***0.33***	***0.44***

Table 6.7: Male/Female bird DNA comparison (sequence 2)

The results of Tables 6.6 and 6.7 indicate partial improvement between the uncorrected and corrected DNA comparisons of potentially different birds. These results illustrate that whilst a 'poor' level of confidence may be placed on the results of an uncorrected comparison, only a 'fair' level of confidence may be obtained from a corrected comparison of these sequences.

6.3.4 Discussion

The results presented illustrate a method of data validation which allows a reliable and repeatable assessment of the quality of a comparison between two DNA sequences. The results of Tables 6.4 and 6.5 indicate significant improvements between corrected and uncorrected assessments of potentially identical DNA sequences. The cross validation results of Tables 6.6 and 6.7 indicate that uncorrected comparisons of potentially different DNA sequences may attain similar assessment results to those obtained from the analysis of uncorrected but potentially identical DNA sequence comparisons. However, the corrected results of the DNA sequence comparisons of Tables 6.4, 6.5, 6.6 and 6.7 indicate that potentially identical DNA sequences attain significantly better validation results than those of potentially different DNA sequences.

6.4 CHAPTER SUMMARY

The results presented in this Chapter indicate the advantages of employing quantitative validation methods within several key application areas. The results from case-study one illustrate the Feature Selective Validation method's ability to not only assess global discrepancies between compared signals, but to classify the nature and magnitudes of these errors. A significant amount of information is provided indicating the magnitudes of error caused by positional differences between trends and features within a comparison, whilst a valid assessment may be obtained indicating structural differences between signals. This type of analysis may be employed to assess the repeatability of results from different test facilities or to assess the quality of experimental procedures between different test personnel.

The results presented in case-study two illustrate the potential advantage of estimating an optimum discretisation level for simulation methods, based on several computational and modelling parameters. This type of analysis may be applied to gain valid arguments as to why further discretisation is necessary, or the possible increase in simulation speed if the level of discretisation within a model is decreased.

The final case-study in this Chapter details the potential advantages of applying the FSV method to DNA sequence comparisons. Employment of the FSC method removes distortions caused by the inconsistent extraction of DNA sequences, whilst the results indicate that only sequences of similar inherent characteristics will benefit from the removal distortions between compared signals.

The FSV method has been applied to three key areas of data analysis, with results illustrating the methods ability to compare data from several application areas employing a universal scaling methodology. The adoption of a clearly defined universal scale allows comparison results from several application areas to be analysed in a quantitative manner with the results from one application area being directly related to those from another. This type of universal validation enhances the information obtained from full systems removing the historical problem of isolation among results taken from different areas of study. For example, within the field of Automotive EMC, it may be applicable to compare validation results obtained from the radiation patterns of mobile communication systems in a vehicle, and biometric validation results obtained from a phantom within that vehicle. Within this universal validation scheme an immense amount of information may be obtained, and both new and existing technologies may be viewed with increased confidence.

CHAPTER 7

DISCUSSION

7. DISCUSSION

Within this project, the underlying mechanisms involved in a visual evaluation of complex data signals have been investigated. Three automated validation procedures have been studied and their potential advantages and disadvantages analysed. From this research a method of quantitative data validation has been developed which overcomes many of the problems inherent in the method of visual evaluation and the past attempts at automated data validation.

7.1 CURRENT AUTOMATED VALIDATION METHODS

Visual evaluation along with three modern methods of automated data validation have been investigated, with their advantages and disadvantages illustrating both the potential pitfalls and benefits of transferring the burden of data validation tasks from man to machine. Conclusions drawn from these methods are summarised below:

7.1.1 Visual evaluation

An in-depth investigation of the most common form of data validation - visual evaluation - has illustrated the complexities involved in developing automated validation methods of equal flexibility. Variabilities between the results of subjects performing visual evaluations are associated with an individuals own paradigms or mental maps. Furthermore, variabilities are inherent between a subjects own assessments for different exposures to a single stimulus. This subjectivity, along with the sheer power of the human visual/perceptual system to abstract information from a stimulus' form makes the process of automating the method of visual evaluation a complicated and volatile procedure.

7.1.2 Correlation

Correlation employs a measure of similarity between two or more data sets, with the results extracted from an assessment indicating a measure of global or overall similarity between the data signals under investigation. However, diagnostic data describing discrepancies at discrete points within the signal spectra's is not attained. This lack of in-depth information regarding the quality of a comparison adversely affects the performance of correlation in situations where powerful diagnostic interpretations are required. Furthermore, whilst the single value extracted from the correlation method is an adequate guide to the nature of a comparison, further information is required to strengthen both the meaning and uniqueness of this measurement. The correlation method is rigid with no flexibility in terms of the measures - atomic, relational and positional differences - taken into account during a visual evaluation of results. Correlation does not provide sufficient information on the quality of a comparison and whilst correlation values obtained from comparisons may be adequate for preliminary assessments of discrepancies, they do not provide the same level of confidence associated with the combined results of subjects performing visual evaluations.

7.1.3 Zanazzi Jona - reliability factor

The reliability factor developed by Zanazzi and Jona[Zanazzi 1977] provides a single value describing the global or overall differences between two or more data signals. This type of analysis offers the potential benefit of providing in-depth diagnostic data at each instantaneous sample within the signal spectra's under investigation. First and second derivatives are employed to emphasis both the shapes and positions of trends and features embedded in the compared signals. Whilst normalisation factors are incorporated into the method to remove the relationship between validation results and the intensities inherent in the signals under investigation. However, discrepancies between amplitude levels are not analysed and a measure of trend and feature shapes is assessed exclusively. This adversely affects the relationship between results extracted from the Zanazzi Jona reliability factor and those from the combined evaluations of subjects participating in identical visual evaluation tasks. Whilst reliability factors provide the potential benefit of lending themselves to the extraction of diagnostic information, Zanazzi and Jona did not develop the method to accommodate this type of in-depth analysis. Furthermore, modifications to the Zanazzi Jona method are complicated due to the complex nature of the algorithm employed to extract validation results. The reliability factor of Zanazzi and Jona illustrates the potential benefits of difference measures over the employment of similarity measures (correlation) employed to analyse complex data signals. However, the method fails to provide in-depth diagnostic information about the quality of a comparison, and values extracted from the method do not lend themselves to straightforward interpretations in terms of the categories employed by humans undertaking visual evaluations.

7.1.4 Van Hove - reliability factor

The reliability factor developed by van Hove *et al*[van Hove 1997] indicates a significant improvement over both correlation and Zanazzi Jona in terms of providing in-depth diagnostic information and individual measurements related to the system of visual evaluation. The method of van Hove employs five individual measures of difference between compared data signals. These five individual reliability factors may be used separately or combined employing weighting factors to form a single global reliability factor. Furthermore, whilst van Hove did not develop the reliability factors to provide in-depth diagnostic information, modifications to the individual algorithms embedded in the method have allowed for the provision of such data. However, the five measurements employed in an analysis of discrepancies between compared signals are only related to two measurements taken into account during a visual evaluation of results. The individual algorithms may be viewed as measurement pairs providing information on amplitude differences and low level trends. Discrepancies between higher order feature shapes/position which relate to a measure of intricate details within the method of visual evaluation are not analysed. This lack of high level information adversely affects the comparison between van Hove reliability factor results and the combined visual evaluations of human subjects. It may be concluded that the van Hove reliability factor provides useful information, however, whilst the structure of an assessment is almost ideal in terms of both isolating discrepancy measurements and employing difference measures to gain diagnostic information, the results do not directly mirror the results obtained from a skilled human performing visual evaluations.

7.2 THE FEATURE SELECTIVE VALIDATION (FSV) METHOD

Results illustrated in Section 5.3 and 5.4 illustrate the ability of the FSV method to mirror the combined visual evaluation results of highly skilled subjects performing identical validation tasks. These results strengthen the hypothesis conjectured in Section 2.2, that the three visual measurements employed by humans comparing data sets may be approximated by a series of absolute, first and second order derivative differences. Furthermore, the accuracy of the FSV results presented in Section 5.3, indicate that whilst a quantitative validation method must produce distinct levels of information based on amplitude levels/positions and feature shapes/positions, a balanced measure of the overall quality of a comparison must be derived from these components.

The results presented in the three case-studies detailed in Sections 6.1, 6.2 and 6.3, illustrate the ability of the FSV method to produce information which is directly related to the classifications employed by humans engaged in visual evaluation studies. Whilst the results of case-study 1 illustrate the potential benefits of classifying errors between two data sets in terms of structural and positional discrepancies. Through the employment of the FSV method, an enormous amount of information in terms of both the methods employed to produce data sets, and the application area in which those methods are used may be obtained. It is to this end that the FSV method may be employed to provide a measure of confidence in the results obtained from new technologies, or procedures.

CHAPTER 8

CONCLUSIONS AND FURTHER WORK

8. CONCLUSIONS AND FURTHER WORK

8.1 CONCLUSIONS

A method of quantitative automatic data validation has been developed. The FSV method[Williams 1998,1999] mirrors the information obtainable from a visual evaluation of results, producing single figures of merit based on discrepancies between amplitude levels and feature shapes/positions. The method was developed to produce a significant amount of diagnostic information in the form of both confidence levels and discrete analyses of compared data sets. Results employing the FSV method have been compared to a significant amount of feedback from engineers and scientists involved in the area of visual evaluation. These results illustrate the FSV method's ability to replicate information produced by the combined assessments of subjects performing identical visual evaluation tasks.

Past methods of automatic validation along with the system of visual evaluation have been researched and their advantages and disadvantages employed to direct the development of the FSV method. The measures employed during an evaluation of two complex signals being based on the mechanisms employed within the human visual/perceptual system. The development of single figures of merit within the FSV method are based on research in the area of correlation. Whilst the development of diagnostic analyses are extracted from the work of van Hove along with considerable research into the benefits of isolating homogeneous regions of signal spectra's before an analysis of discrepancies is employed. The seed for emphasis routines embedded in the FSV method were taken from the area of Reliability factors, however considerable modification to these emphasis algorithms was necessary before they could be incorporated as an integral part of the validation method.

The inclusion of variable weighting factors (objective subjectivity) associated with the two main measurements (amplitudes and features) within the FSV method have allowed the validation of results from a wide cross section of application areas. Past methods of automated data validation have relied on rigid algorithms which do not lend themselves to the same flexibility inherent in the process of human visual evaluation. Subjectivity within humans allows validation of diverse data sets with the information gained conforming to a single global interpretation scale. The FSV method embraces this type of 'objective subjectivity', and it is this flexibility which allows a wide cross section of data sets to be validated, with the results from one area of study being directly comparable to those validated from an entirely different area.

Employing the FSV method and more specifically the GDM value, a method of data signal correction was developed to align distorted signals. This method has been invaluable in areas such as DNA fingerprint analysis where results employing different gels incur considerable distortions. Furthermore, the employment of the FSC method allows the classification of errors in terms of both position and structure.

The project aims set out in Chapter 1 have been met through the development of a flexible, quantitative validation method which allows users to automatically assess complex data sets in a consistent manner. The FSV method may be employed to: determine optimum resolutions for numerical models[Williams 1997]; assess the magnitude of repeatability inherent in experimental equipment and/or procedures[Duffy 1998, Ruddle 1998,1999]; or correct and analyse distorted signal sets[Williams 1999]. The flexible framework within which the FSV method operates, allows simple pass/fail analyses of compared signals based on single figures of merit, and the provision of associated confidence levels for a comparison. Whilst diagnostic information is made available for a rigorous and focused analysis into the nature, location and magnitude of errors impinging on a comparison of two or more complex signals.

8.2 FURTHER WORK

Further work should concentrate on extending the capabilities of the FSV method. The validation of results with two or more dimensions (e.g. surface and volumetric) should be investigated, and the capabilities of the FSV method should be extended to incorporate this type of advanced analysis.

The FSV and FSC methods should be optimised through the employment of Parallel Virtual Machine (PVM) software[Geist 1994, PVM 1999]. Through network connections, a parallel machine may be built employing distributed processing software. In this way, computation may be distributed between several independent numerical platforms. Initial studies (detailed in Section 4.3.4) in this area have shown that whilst the speed of FSC method may be increased, several levels of optimisation are at present still required. Fundamentally, the parallel FSC method requires a dynamic or adaptive computational management system, which schedules the parallel processes within the scheme. This dynamic process may be based on heuristics (rule based) or learning algorithms such as genetic algorithms. However in both cases, the management system must possess the ability to change its actions (through feedback) to compensate for the changing states (varying load averages, loss of network connections, etc.) of the hardware on which it is employed.

Enhancements to the quality of feedback from the FSV method may be sort through the employment of higher level analyses such as neural networks or genetic algorithms. Allowing intelligent computational judgements to be made on the quality of compared signals and the direct optimisation of data acquisition methods (e.g. the automated redesign of numerical models).

CHAPTER 9

REFERENCES

9. REFERENCES

ALPHA, 1999, "High Powered Computing", **http://www.digital.com/info/hpc/hpc.html**.

Ames LB, Learned J, Metraux R and Walker R, 1953, "Development of Perception in the Young Child as Observed in Response to Rorschach Test Blots", **Journal of Genetic Psychology**, 82, pp 183.

Balazs I, Baird M, McElfresh K and Shaler R, 1990, "Experimental Techniques for the Isolation and Analysis of DNA from Forensic Materials", in Robertson J, Ross AM and Burgoyne LA (Editor), DNA in Forensic Science (**Ellis Horwood, Chichester**).

Beowulf, 1999, "Parallel Programming Clusters", **http://www.beowulf.org/**

Berlyne DE, 1960, "Conflict, Arousal and Curiosity", (**McGraw-Hill, London**).

Booton RC, jr., 1992, "Computational Methods for Electromagnetics and Microwaves", (**Wiley, London**).

Carmichael L, Hogan HP and Walter AA, 1932, "An Experimental study of the Effect of Language on the Reproduction of Visually Perceived Form" , **Journal of Experimental Psychology**, 15, pp 73.

Christopoulos C., 1995, "The Transmission-Line Method TLM", (**Oxford University Press**).

Clarke LJ, 1985, "Surface Crystallography", (**Wiley, London**).

Cook WM, 1931, "Ability of Children in Colour Discrimination", **Child Development**, 2, pp 303.

Covey SR, 1994, "The Seven Habits of Highly Effective People", (**Simon and Schuster, London**).

December WN, 1960, "The Psychology of Perception", **(Holt, New York)**.

Duffy AP, Benson TM and Christopoulos C, 1995, "Design and Analysis of Dielectric Loaded Cavity Backed Aperture Antennas using Transmission Line Modelling (TLM)", **9th International Conference on Antennas and Propagation**, 1, pp 21.

Duffy AP, Woolfson MS and Benson TM, 1994, "Use of Correlation Functions to Assist the Experimental Validation of Numerical Modelling Techniques", **Microwave and Optical Technology Letters**, 7(8), pp 361.

Duffy AP, Williams AJM, Benson TM and Woolfson MS, 1998, "A Powerful Technique for the Quantitative Assessment of Experimental Repeatability", **IEE Proceedings on Science Measurement and Technology**, 145(4), pp 177.

Eriksen CW, 1990, "Attentional Search of the Visual Field", in Brogan D (Editor), Visual Search **(Taylor and Francis, London)**.

Fernandez DP and Alberola CR, 1990, "Do Visual Search Variables Affect Memory Search", in Brogan D (Editor), Visual Search **(Taylor and Francis, London)**.

Fozard JL, Wolf E, Bell B, McFarland RA and Podolsky S, 1977, "Visual Perception and Communication", in Birren JE and Schale KW (Editors) Handbook of Psychology of Ageing **(von Nostrand, New York)**.

Geist A, Beguelin A, Dongarra J, Jiang W, Mancheck R and Sunderam V, 1994, "PVM: Parallel Virtual Machine - A Users' Guide and Tutorial for Networked Parallel Computing", **(Mitt-Press)**.

Ghent L, 1956, "Perception of Overlapping and Embedded Figures by Children of Different Ages", **American Journal of Psychology**, 69, pp 575.

Granit AR, 1921, "A Study of the Perception of Form", **British Journal of Psychology**, 12, pp 223.

Haith MM, Morrison FJ and Sheingold K, 1970, "Tachistoscopic Recognition of Geometric Forms by Children and Adults", **Psychonometric Science**, 19, pp 345.

Herman DT, Lawless RH and Marshall RW, 1957, "Variables in the Effect of Language on the Reproduction of Visually Perceived Forms", **Perceptual Motor Skills**, 7, pp 171.

Hilsenrath OA and Zeevi YY, 1990, "Feature Extraction and Sensitivity Matching in Visual Search in Man and Machine", in Brogan D (Editor), Visual Search (**Taylor and Francis, London**).

Johns DP, Scaramuzza RA and Wlodarczyk AJ, 1995, "Micro-stripes - Microwave Design Tool Based on 3D-TLM", **First International Workshop on TLM University of Victoria,** pp 169.

Johnson J and Picton P, 1995, "Concepts in Artificial Intelligence: Volume II", (**Butterworth and Heinemann, London**).

Kincaid D and Cheney W, 1991, "Numerical Analysis", (**Brooks and Cole**).

Koffa K and Harrower M, 1931, "Colour and Organisation", **Psychology Forsch.**, 15, pp 193.

Koffa K, 1935, "Principles of Gestalt Psychology", (**Kegan Paul, London**).

Kristofferson AB, 1957, "Visual Detection as Influenced by Target Form", in Wulfeck JW and Taylor JH (Editors), Form Discrimination as related to Military Problems (**National Academy of Science, Washington D.C.**).

Krose BJA and Julesz B, 1990, "Automatic or Voluntary Allocation of Attention in Visual Search Tasks", in Brogan D (Editor), Visual Search (**Taylor and Francis, London**).

Lindsley DB, 1957, "The Reticular System and Perceptual Discrimination", in Jasper HH et al. (Editors), Reticular Formation of the Brain (**Churchill, London**).

Mattison PE, 1994, "Practical Digital Video with Programming Examples in C", **(Wiley, London)**

Menacer M, Woolfson MS and Crowe JA, 1992, "Method of Correction for Linear-Distortion of Signals and Application to DNA Fingerprint Data", **Electronics Letters**, 28, pp 2126.

Neisser U, 1970, "Visual Search", **Scientific American**, 210(6), pp 94.

Norton D and Stark L, 1971a, "Scanpaths in Eye Movements During Pattern Recognition", **Science**, 171, pp 308.

Norton D and Stark L, 1971b, "Scanpaths in Saccadic Eye Movements While Viewing and Recognising Patterns", **Vision Research**, 11, pp 928.

Pendry JB, 1980, "Reliability Factors for LEED Calculations", **Journal of Physics C: Solid State Physics**, 13, pp 937.

Postman N, 1992, "Technopoly", (**Knopf, New York**).

Purrello M and Balazs I, 1983, "Direct Hybridisation of Single-Copy DNA Sequences to DNA in Agarose Gels", (**Anal. Biochem.**), 128, pp 503.

PVM, 1999, "Parallel Virtual Machine", **http://www.epm.ornl.gov/pvm/**.

Ruddle A, Ward D, Williams AJM and Duffy AP, 1998, "Objective Validation of Automotive EMC Models", **IEEE International Symposium on EMC, Denver**, 4, pp 20.

Ruddle A, Williams AJMW, Ward D and Duffy AP, 1999, "Objective Validation Techniques for Automotive EMC Engineering", **IEE Colloquium on Electromagnetic Compatibility for Automotive Electronics**.

Schneider W, and Shiffrin RM, 1977, "Controlled and Automatic Human Information Processing: I. Detection, Search and Attention", **Psychological Review**, 34, pp 1.

Shiffrin RM, and Schneider W, 1977, "Controlled and Automatic Human Information Processing: II. Perceptual learning, Automatic Attending, and General Theory", **Psychological Review**, 34, pp 129.

Terman LM, 1937, "Measuring Intelligence", (**Harrap, London**).

Usher MJ, 1984, "Information Theory for Information Technologists", (**Macmillan**).

Van Hove MA, Tong SY and Elconin MH, 1997, "Surface Structure Refinements of 2H-MoS_2, 2H-$NbSe_2$ and W(100)p(2 × 1)-O Via New Reliability Factors for Surface Crystallography", **Surface Science**, 64, pp 85.

Venables PH and Warwick-Evens LA, 1967, "Cortical Arousal and Two Flash Thresholds", **Psychonometric Science**, 8, pp 231.

Vernon M D, 1975, "The Psychology of Perception", (**Penguin Books**).

Westcott MR and Tolchin M, 1968, "Individual and Age Related Differences in Perceptual Inference Behaviour", **Perceptual Motor Skills**, 26, pp 683.

Williams AJM, Duffy AP, Woolfson MS and Benson TM, 1997, "A Novel Technique for Quantifying Model Refinements as a Function of Computational Resources", **Applied Computational Electromagnetics Society, UK Workshop**.

Williams AJM, Duffy AP, Woolfson MS and Benson TM, 1998, "A Powerful New Technique for the Quantitative Validation of Modelled and Experimental Data", **Microwave and Optical Technology Letters**, 17(5), pp 284.

Williams AJM and Duffy AP, 1999, "The Quantitative Validation Project (QVP)", **http://www.eng.dmu.ac.uk/fsv/**.

Witkin HA, Lewis HB, Hertzman M, Machover K, Meissener PB and Wapner S, 1954, "Personality Through Perception", (**Harper, New York**).

Woolfson MS, Benson TM, Christopoulos C and Duffy AP, 1995, "Quantitative Assessment of the Comparison of Electromagnetic Calculations with Experimental Data", **Applied Computational Electromagnetics Society Newsletter**, 10(1), pp 34.

Yarbus AL, 1967, "Eye Movements and Vision", (**Plenum Press, New York**).

Zanazzi E and Jona F, 1977, "Reliability Factor for Surface Structure Determinations", **Surface Science**, 62, pp 61.

CHAPTER 10

PUBLICATIONS

10. PUBLICATIONS

10.1 PUBLISHED PAPERS

Williams AJM, Benson TM and Duffy AP, 1997, "Determining the Accuracy of TLM in Simulating the Behaviour of Resonant Cavities with Arbitrary Dielectric Loading", **International Journal of Electronics**, 83(5), pp 645.

Williams AJM, Woolfson MS, Benson TM and Duffy AP, 1997, "Quantitative Validation of Numerical Technique Results against Experimental Data", **IEE 10th International Conference on Antennas and Propagation, York**, 1, pp 1.532.

Williams AJM, Duffy AP and Scaramuzza RA, 1997, "A Modelling Approach to Determining the Effective Working Volume of a Mode Stirred Chamber", **IEE 10th International conference on Electromagnetic Compatibility, Warwick**, pp 187.

Williams AJM, Duffy AP, Woolfson MS and Benson TM, 1997, "A Novel Technique for Quantifying Model Refinements as a Function of Computational Resources", **Applied Computational Electromagnetics Society (UK Workshop), London**.

Duffy AP, Williams AJM, Benson TM and Woolfson MS, 1997, "A Powerful Technique for the Quantitative Assessment of Experimental repeatability", **8th International Conference on Electromagnetic Measurement, London,** pp 38.1.

Williams AJM, Duffy AP, Woolfson MS and Benson TM, 1998, "A Powerful New Technique for the Quantitative Validation of Modelled and Experimental Data", **Microwave and Optical Technology Letters**, 17(5), pp 284.

Duffy AP, Williams AJM, Benson TM and Woolfson MS, 1998, "A Powerful Technique for the Quantitative Assessment of Experimental Repeatability", **IEE Proceedings on Science Measurement and Technology**, 145(4), pp 177.

Ruddle A, Ward D, Williams AJM and Duffy AP, 1998, "Objective Validation of Automotive EMC Models", **IEEE International Symposium on EMC, Denver,** 4, pp 20.

Duffy AP and Williams AJM, 1999, "Optimising Mode Stirred Chambers", **13th International Symposium on EMC**, pp 685.

Ruddle A, Williams AJM, Ward D and Duffy AP, 1999, "Objective Validation Techniques for Automotive EMC Engineering", **IEE Colloquium on Electromagnetic Compatibility for Automotive Electronics**.